DOCUMENT RESUME

ED 044 957 48 FL 001 967

AUTHOR	Paquette, F. Andre; Tollinger, Suzanne
TITLE	A Handbook on Foreign Language Classroom Testing: French: German, Italian, Russian, Spanish.
INSTITUTION	Modern Language Association of America, New York, N.Y.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO	BR-6-2619
PUB DATE	Jun 68
CONTRACT	OEC-1-6-062619-1876
NOTE	234p.
AVAILABLE FROM	MLA Materials Center, 62 Fifth Avenue, New York, N.Y. 10011 ($6.00)
EDRS PRICE	EDRS Price MF-$1.00 HC-$11.80
DESCRIPTORS	Achievement Tests, Annotated Bibliographies, French, German, Instructional Program Divisions, Italian, Item Analysis, *Language Instruction, *Language Tests, *Modern Languages, *Objective Tests, Russian, Spanish, Statistical Analysis, Teaching Guides, *Testing, Test Interpretation

ABSTRACT

This handbook illustrates how the classroom teacher can make efficient use of tests, discusses the main purposes tests can serve, treats principal kinds of testing devices especially useful in language instruction, and includes a section on the interpretation of test results. Chapters include: (1) the importance and place of testing in the foreign language program, (2) planning the classroom test, (3) construction of test items, (4) preparing test items in French, German, Italian, Russian, and Spanish, and (5) test interpretation. A glossary of technical terms used in the handbook and an annotated bibliography on testing are included. (RL)

OEC-1-6-062619-1876
TITLE VI, NDEA, Section 602
PA-48

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED044957

FL 001 967

A HANDBOOK ON FOREIGN LANGUAGE CLASSROOM TESTING:

FRENCH, GERMAN, ITALIAN, RUSSIAN, SPANISH

June 1968

by

MODERN LANGUAGE ASSOCIATION OF AMERICA

F. André Paquette, Project Director
Suzanne Tollinger, Research Assistant

The research reported herein was performed pursuant to a contract with the Office of Education, United States Department of Health, Education, and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

PREFACE

This *Handbook* reflects the efforts of many people in testing and foreign language teaching. The participants in the conferences to organize the writing of the *Handbook* included foreign language teachers at the elementary, high school, and college levels, foreign language methods teachers, and measurements specialists:

Richard Barrutia
University of California
at Irvine

Frederick Bosco
Catholic University of America
Washington, D. C.

Ludmilla Bradley
Fullerton Junior College
Fullerton, Calif.

Nancy Capozzi
Public School 41 R
Staten Island, N. Y.

Guillermo del Olmo
Rutgers University
New Brunswick, N. J.

Robert J. Di Pietro
Georgetown University
Washington, D. C.

Elisabeth Epting
Converse College
Spartanburg, S. C.

Elizabeth A. Gorra
Eastern Junior High School
Greenwich, Conn.

Mary Heiser
University of Southern California
at Los Angeles

Carlene Horvath
Parkland Joint School District
Orefield, Pa.

Albert JeKenta
Beverly Hills Unified School
District
Beverly Hills, Calif.

Tom Kelly
Parkland High School
Orefield, Pa.

Robert L. Lathrop
Pennsylvania State University
University Park, Pa.

Louise Lillard
Beverly Hills High School
Beverly Hills, Calif.

Klaus A. Mueller
University of California
at Berkeley

Josefina V. O'Keefe
Bellevue School District
Bellevue, Washington

Elizabeth Ratté
Purdue University
Lafayette, Ind.

Betty A. Robertson
Northport High School
Northport, N. Y.

Dan Romani
Southern Illinois University
Edwardsville, Ill.

Lorraine Strasheim
Indiana University
Bloomington, Ind.

Robert Stecklein
Bureau of Institutional Research
University of Minnesota
Minneapolis, Minn.

Allan Taylor
University of Colorado
Boulder, Colo.

The idea for this project emanated from the discussions of the Test Advisory Committee, under whose advisorship the MLA Testing Program was conducted:

James Alatis
Georgetown University
Washington, D. C.

Don R. Iodice
Oakland University
Rochester, Mich.

Albert JeKenta
Beverly Hills Unified
School District
Beverly Hills, Calif.

Albert H. Marckwardt
Princeton University
Princeton, N. J.

Sanford Newell
Converse College
Spartanburg, S. C.

Sol Saporta
University of Washington
Seattle, Wash.

Irving Wershow
University of Florida
Gainesville, Fla.

We would like to thank the following for permission to use copyright materials:

Modern Language Association of America, for the permission to use several definitions from What's What: A List of Useful Terms for the Teacher of Modern Languages, compiled by Donald D. Walsh, Copyright MLA 1963, 1964, 1965.

John E. Stecklein, Director, Bureau of Institutional Research, University of Minnesota, for permission to use excerpts from Bulletins on Classroom Testing, Numbers 2, 3, 4, 5, 6, and 7, Bureau of Institutional Research, University of Minnesota, Copyright 1954, 1955, 1956, University of Minnesota. This material has been recombined to form Chapter III of this Handbook. A complete set of these booklets may be obtained by writing to the Bureau of Institutional Research, 3338 University Avenue, S. E., Minneapolis, Minn. 55415.

At various times during the development of this Handbook, most of the sample test items were reviewed by the Foreign Language Test Staff of Educational Testing Service, Princeton, N. J. 08540.

CONTENTS

Preface . . . . Page iii

Introduction ix

Chapter I The Importance and Place of Testing in the Foreign-Language Program 1

Chapter II Planning the Classroom Test 12

Chapter III The Construction of Test Items 28

Chapter IV Preparing Test Items

French 40

German 67

Italian 89

Russian 116

Spanish 144

Chapter V The Interpretation and Use of Test Results 173

Glossary 208

Bibliography 211

INTRODUCTION

This Handbook has been produced to help the classroom teacher make efficient use of tests. It contains discussions of the main purposes tests can serve. It treats the principal kinds of testing devices that seem to be especially useful for those purposes in the context of foreign-language teaching. And it has a section (Chapter V) on the interpretation of test results.

Much of what is contained in this Handbook is useful at all levels of instruction and with people of all ages. Some testing devices, however, are obviously not appropriate to children in elementary or even junior-high-school classes. The teacher should have no difficulty making a judicious choice.

While most teachers have had some formal instruction in their own undergraduate education in the general principles of measurement, very few classroom teachers seem to have had any formal instruction in principles and methods of testing in the specific context of foreign-language teaching. Much of the general theoretical discussion contained in this Handbook will therefore be more or less familiar to the reader. It is included here for several reasons: to make the Handbook a self-contained unit; to remind the reader of notions he may have forgotten; and to fill in gaps in the reader's background. The reader will notice that certain topics are treated in several different chapters. This repetition is deliberate. In each case, one treatment is much more thorough than the others; these latter are intended to serve as useful contexts for discussion of other topics. A Handbook is, after all, a reference work: the teacher should find at one place therein whatever he needs to know about the topic he has looked up.

This Handbook is not complete. No book of this kind can be complete. It may suffice for many teachers. It may inspire others to exercise their imagination and ingenuity. It may lead still others to work through more technical expositions of some of the topics treated here. Thos who have helped to prepare this Handbook hope it will be useful in some way to all of their colleagues.

Chapter I

THE IMPORTANCE AND PLACE OF TESTING IN THE FOREIGN-LANGUAGE PROGRAM

1.1 USES OF TESTS

1.1.1 Knowledge and skills can be tested in many ways and for many reasons. By its very nature audio-lingual teaching in particular requires a constant flow of information and encouragement of the kinds supplied by regular and varied testing. It is useful to consider tests as (1) measures of achievement, as (2) diagnostic instruments, and as (3) teaching tools. The preparation of adequate instruments to measure aptitude requires much time and effort of highly qualified specialists. Teachers should not try to prepare their own, since carefully prepared tests to measure language aptitude are readily available.[1]

1.1.2 Measuring Achievement

One of the most common functions that tests serve is to measure a student's achievement. Achievement tests establish for the teacher a carefully selected sample of a student's performance.

1.1.3 Tests as Diagnostic Instruments

Tests constructed for diagnostic purposes are different from tests designed specifically to measure language aptitude. Diagnostic tests meet a continuing everyday need and are helpful to the teacher throughout the course of instruction. From the results of a test, a teacher can diagnose possible inadequacies in either his teaching or his teaching materials. Careful inspection of an individual student's paper will also often make clear what specific difficulties he is having. The teacher can use such knowledge to improve his teaching and the student's learning.

---

[1] John B. Carroll and Stanley M. Sapon, Modern Language Aptitude Test, Form A (New York: The Psychological Corporation, 1958, 1959).

Pimsleur Language Aptitude Battery (New York: Harcourt, Brace & World, Inc., 1966).

1.1.4 Tests as Teaching Tools

It can be argued that students can actually learn in the process of taking a test. Such is likely to be the case if the test challenges the student to learn by analogy and discovery and leads him to try to understand new notions as the test proceeds. Such tests not only test what the student has learned. They help the teacher to decide whether the student is learning how to learn language. Some items can be included in each test for just this purpose; they may be very helpful for student and teacher alike in carrying on the foreign-language program.

1.1.5 Collateral Problems and Benefits

1.1.5.1 The best tests serve several purposes simultaneously. Classroom tests can provide a basis for assigning grades, for comparing pupils with one another, for motivating learning, for directing learning, and for helping to improve instruction in other ways. Classroom tests also serve students as a source of information about the objectives of the course and their own individual achievement. If a test faithfully represents the major objectives of a course, preparing for it and reviewing it after it has been administered will reinforce the rest of the work of the course.

1.1.5.2 Testing also often confirms the evaluation which a good teacher has already made of a student. If a test does not confirm the teacher's opinion, several explanations are possible. The test itself may be badly constructed; a consistent pattern of low scores for good students or high scores for weak ones, for instance, usually means that the test itself is defective. The teacher's previous appraisal may have been based on insufficient information. The item types chosen may not have been used often enough previously for the students to have become familiar with them; they may therefore have misunderstood what was expected of them and have given incorrect answers, even though they knew the correct answers. A teacher's day-to-day evaluation of a student may be quite correct, but formal testing required primarily to give the carefully measured results necessary for permanent records may not be enough to corroborate that appraisal.

1.1.5.3 The questions used in classroom tests provide students with information about the aims of the course. They also direct the students' efforts towards the acquisition of the specific facts and skills needed to achieve those aims. When the teacher sets test specifications and devises test questions to help him to judge his students' progress toward particular goals, he may realize that those goals have implications of which he was not aware.

1.1.6 Spaced Testing

The principle of spaced learning is that in foreign-language teaching it is better to teach or drill a given item for two minutes on ten different occasions than for twenty minutes at one time. How is this principle related to testing? More frequent short tests seem generally to be more effective than infrequent long tests. There are a number of reasons why frequent short tests are useful. If students gradually become used to a variety of different types of items on short tests, longer, more complex examinations are more just and more reliable because the students are comfortable with their formats. By using frequent examinations, the teacher may diagnose problems he can correct as he goes along. Frequent tests also insure that most students will study regularly and diligently.

1.2 SPECIFIC TECHNIQUES

1.2.1 To determine whether students can make logical and well-chosen changes in the dialogues they memorize, a simple method is to put on the blackboard a series of phrases from which they are to select in order to vary the dialogue. Pairs of students are asked to make substitutions orally by selecting appropriate phrases from among those listed on the blackboard. Each partner is obliged to adjust to the accumulated variations and, in turn, to make substitutions that fit the rest of the evolving conversation. With this technique, the teacher not only tests the student's ability to read aloud but his ability to understand what is said to him. Such a test teaches at the same time as it tests.

1.2.2 Impromptu Testing

The format of an oral test need not be rigid. The seemingly unstructured nature of a test may be its best quality. For example, the teacher may simply move about the room quickly, asking as many short, quick questions as he can in ten minutes. These questions can be varied but must stay, of course, within the student's range of vocabulary and grammar. One problem which arises with this procedure is that there is not enough time to grade the student's responses. This difficulty can be overcome with a pre-taped test which will be described below (see 1.2.3.2).

1.2.3 Directed Responses

1.2.3.1 Another type of loosely structured oral test derives from directed dialogue. The teacher instructs Student One to tell Student Two to ask Student Three whether he has done, or will do, something for him (Student One). It is important that all three students become engaged in the exchange so that they all have to modify answers and inflect verbs in various ways. Here, too, testing and teaching are not easily separated.

1.2.3.2 Both of the techniques described in 1.2.2 and 1.2.3.1 have the disadvantage of being difficult to grade because they happen so quickly and because the teacher must therefore rely on his memory when he records his evaluation of each student's performance. The following arrangement largely overcomes this disadvantage. Pre-record a large number of questions and statements that permit rejoinders. Leave a pause after each question or statement long enough for a student to answer, then record a random number ("one", "two", etc.--but in the foreign language) followed by another pause. Make as many such items as necessary to test adequately all of the students in the class. Assign numbers to the students especially for the test. Play the tape, and have each student answer whenever his number comes up; he must also write out his answer on the test sheet. The students will be alert throughout the test since the numbers are recorded at random. They will all formulate answers to all of the questions, even though each individual student will actually have to answer only five or six questions. The teacher is free to move about with grade book in hand to make judgments and record grades. The student has enough time to write out his answers, because no number is ever recorded twice in succession. The test papers can be collected and used by the teacher to decide on an additional grade for writing.

1.2.4 Pattern Paragraph Practice

Students are better able to handle test items the more they resemble in format the regular classroom presentation of the material being tested. Items of a type familiar to the student have much higher reliability than those that present material in a format the student has never encountered. Suppose a class has been practicing paragraph writing by replacing lexical items (nouns, verbs, etc.) while leaving function words (prepositions, conjunctions, etc.) intact, thereby adhering to the original grammatical structure of the paragraph but altering its content. For practice in writing, the instructions could be: "Copy the following short paragraph first as it is. Then rewrite it changing the underlined words to words of your own choice that make sense and are grammatically correct within the reconstructed whole." For test purposes: "Rewrite the following paragraph, changing each underlined word to any appropriate word of your choice."

1.3 MOTIVATION

In addition to verifying assimilation of the material covered, good tests motivate students to strive to attain certain objectives. When he is taking a test, the student is at the peak of his powers of attention and concentration. Some would argue that the teacher should take advantage of this favorable situation and include some test items which are clearly designed to lead the student to see new relationships--which will lead him to learn something new. More generally, it is clear that much of a student's best studying is done when he is preparing for a quiz or an examination. This heightened and sharpened effort can be focused directly on whatever the teacher chooses to emphasize. The teacher need only be sure that the student knows what to expect on the test: what material will be covered, and what testing techniques will be used. Testing is one of the strongest motivating forces available to the teacher; it is a force that can be focused clearly on specific objectives. Short, frequent tests concentrating on one or two skills at a time will tend to sustain a relatively high level of concentration.

1.4 GENERALIZATIONS

If the student does not know in advance exactly which uses of the subjunctive, for example, will appear on a test, he will tend to study the whole category. In a test of this type, however, the reliability of the test might be open to question if it is too short or if there are not enough items (approximately three) testing each specific point. (Pairs of items permit more reliable judgments than single items. Sets of three items are still more reliable than pairs.) For the most part it is safe to say that if a student can conjugate properly three regular verbs, he can conjugate any number of others, that if he can write two correct contrary-to-fact sentences, he can write another, etc. It also seems reasonable to conclude that if a student can write an acceptable paragraph, he could write an acceptable composition. The consistency of performance of a group of students also allows us to generalize about the rank order of the group: the best student in the class on one day is not likely to be the worst on the next.

1.5 LISTENING COMPREHENSION

1.5.1 Comprehension practice and tests can be conducted efficiently by having the student listen to recorded material with which he is unfamiliar either in the language laboratory or at home. His task is to write out a given text, taking as much time as he needs to transcribe it all as correctly as he can. In the classroom, the teacher may have several students copy sections of the text onto the blackboard in order to have the entire script readily available for the whole class. The whole piece can then be carefully studied for meaning, grammatical difficulties, and orthography. It can also be read aloud.

1.5.2 A text to be used for a comprehension test may be a pre-recorded paragraph, dialogue, or narration with which the student is expected to deal in almost this same way. He is limited, of course, in the amount of time he can spend writing out what he hears. Such a test is essentially a dictation. (The text can also be read aloud by the teacher.)

1.6 A NOTE OF CAUTION

1.6.1 Teaching and testing are not so much alike that all testing techniques should be used or can be used for teaching. While most teaching techniques can be used for testing, the reverse is not true. For instance, a good test should test everyone in a class. It must challenge and indeed go slightly beyond the competence of the best student in the class; otherwise, the limit of that student's ability remains unknown. If we take this view, even one perfect paper means that the test has not really measured the whole class. But a single student does not invalidate the test as a good measure of the great majority of the students in the class. It is dangerous always to aim over the head of the best student; every student needs positive reinforcement continually, or he may lose interest and momentum. On the other hand, if a test item is so easy that every student answers it correctly, it cannot discriminate between good and poor students. Whenever the tester's purpose is to make such a discrimination, items of this kind are invalid.[2] Those same items, however, may measure very successfully the achievement of a class. Individual test items are valid or invalid as they measure or fail to measure what the examiner wants to evaluate. (See Chapter V.) In class we regularly provide challenges to which the entire class can respond successfully. Such drills teach and simultaneously permit the teacher to evaluate each student's achievement. If they are carefully constructed, however, and if the whole class has learned successfully the point at issue, each student will respond correctly; such techniques are useful for evaluating the students' performances, but they are not very useful for ranking them. This is an example of a teaching technique that can be used for testing for one purpose, but not for another.

1.6.2 The student must know precisely what material a test will cover and what skills will be emphasized, but at the same time he should understand that acquisition of a second language is necessarily a cumulative process and that grammar covered at any given point in a course naturally continues to be tested incidentally from then on. There should be frequent tests deliberately recombining a number of points. Some students who do well when the class is working on a particular topic may be unable to produce those same structures spontaneously at a later time. When this occurs on a test, the test is useful for diagnostic purposes. The teacher should take information of this kind into account in planning his course.

---

[2] Cf. validity: the extent to which a test measures what it is supposed to measure.

1.7 CONCLUSIONS

(a) Tests can be used to measure (1) aptitude, and (2) achievement; they can be used for (3) diagnostic and (4) teaching purposes.

(b) A good test will bear out what the teacher already knows about most students' achievement.

(c) When a test does not confirm a teacher's earlier judgment of a class or a student, it is especially useful as a diagnostic tool.

(d) Tests can sometimes actually teach by imitation, by induction, or by analogy.

(e) The principle of spaced learning applies equally well to testing. Frequent, short tests are more reliable, more productive, and fairer to the students than infrequent long ones. But both shorter and longer, more comprehensive tests should be used.

(f) Many teaching techniques are also excellent for testing.

(g) Short tests are reliable if they contain enough items on each point tested.

(h) Comprehension practice is often neglected.

(i) Tests can be a powerful means with which to motivate students to strive to reach specific objectives.

(j) Grading oral production can be made easier by the use of a simple tape-recorded test.

Chapter II

PLANNING THE CLASSROOM TEST

2.1 WHAT TO TEST

2.1.1 A test is a sample of a student's work. Because it is generally not possible to cover in a test everything that a student has been taught, a test contains a selection from among all of the questions that he might possibly have been asked on specific topics. This sample should enable the teacher to generalize about the student's progress or achievement in the total area from which the sample is drawn. Therefore, before the teacher begins to compose a test, he should have clearly in mind the purpose the test is to serve, the specific language skills and content areas that are to be measured, and the emphasis he wants to put on various aspects of achievement. If these factors are not taken into account in preparing a classroom test, there is a serious risk that the test may provide a dangerously misleading picture of the students' progress.

2.1.2 If the teacher writes and assembles test items without a clear plan, he may overemphasize language skills and content areas in which items are easy to construct, simple to administer, and easy to grade, or in which he himself is especially interested even though they may not reflect accurately what has been taught in class.

2.1.3 Ideally, in planning a test the teacher should specify the language skills and content areas to be measured and the relative emphasis to be assigned them in much the same way as do the publishers of standardized tests. Setting the specifications for a test need not, however, be an especially elaborate process, although for major examinations it is more exacting than for simple quizzes. Sometimes all that is needed is a simple list of topics that must be included to provide a representative sample of the language skills and content areas to be covered in the test together with an indication of their relative importance.

2.1.4 The teacher must decide when to test, what kinds of items to use, how many of each to include, how the test should be given, who should score it, whether or not the scores should be corrected for guessing, and what the student's grade will mean.

2.2 WHEN TO TEST

2.2.1 All tests should serve a useful purpose. Tests should not be given just for the sake of testing; they should not be used as instruments of punishment; and they should not be given to free the teacher to attend to other tasks.

However, the testing schedule is frequently regulated by the school's fixed schedule for marking and reporting grades. Some teachers test only at the end of the marking period. Others give full-period tests every four, six, or eight weeks, or at the completion of major units of study. Teachers who make use of tests for instructional purposes give short tests frequently; for instance, some use unannounced quizzes to motivate students to do their assignments carefully and on time.

2.2.2 Frequent tests have several advantages. Both teacher and student are kept better informed about the student's progress; the teacher has a larger number of observations on which to base grades; the effectiveness of the teaching of specific topics is evaluated before it is too late to correct misconceptions. However, tests can be too frequent; the student can easily be led to expend too much effort preparing for tests for the sake of grades.

2.2.3 The time within the school day when a test can be given is usually predetermined by the school schedule. In any case, there is very little evidence supporting preference for any particular day or hour for testing. It is only sensible, however, to avoid scheduling tests on the morning after an evening affair like a school dance or for the hour following an exciting activity like a football rally.

2.2.4 If possible, it is desirable to schedule end-of-year tests before the closing of school so that the teacher can discuss the test results with the students. Unfortunately, many final-examination schedules do not allow time for students to find out what questions they missed or why they missed them.

2.3 TYPES OF TEST ITEMS

2.3.1 The relative merits of essay questions and objective items have been much debated. Neither type of item is inherently superior to the other. The important questions are (1) whether a type of item serves the purpose of the test, and (2) how well an item is constructed.

2.3.2 Essay Questions

The essay question requires the student to develop answers from his own knowledge without benefit of suggested possibilities, and to express the answer in his own words. When a test is concerned with cultural information, the content of literary works, or literary analysis, essay questions are clearly appropriate. Essay questions can be prepared easily. They can often simply be written on the blackboard. Essay questions very largely eliminate guessing. Essay questions can serve to measure complex abilities when students are required to describe, explain, compare, contrast, analyze, criticize, interpret, or generalize. Unfortunately, however, many essay questions do not measure such abilities. All too often students are asked merely to state, in continuous prose, facts or information that they are supposed to have learned. (At early and intermediate stages of language study, it is wholly appropriate, of course, to ask the student simply to state facts or opinion in straightforward prose. More complex tasks should be reserved for advanced courses.) Essay questions may be stated so ambiguously or in such general terms that it is difficult to decide whether the student has really treated the topic.

2.3.3 The principal disadvantage of the essay question is the unreliability of the scoring of the answer. One reason why it is difficult to achieve reliability in scoring answers to essay questions is that teachers differ greatly in their judgments. Such differences of opinion are usually due to the simple fact that teachers can easily judge a paper very differently if their criteria are very different. The teacher's judgment may even be influenced by how a paper looks--the easier the paper is to read, the higher the grade assigned.[*] The teacher's judgment may be influenced by the "halo" effect: he may tend to mark a student according to preconceived notions of the student's ability or of the general ability of the class as a whole. The "halo" effect may also operate from question to question: the quality of the answer to the first question may affect the scoring of the answers to subsequent questions. Finally, teachers frequently correct tests after a long school day, even late in the evening, and their scoring may therefore be directly affected by real fatigue.

2.3.4 There are two other limitations to essay questions. First, there is the possibility of inadequacy of sampling when only a few items are included in a test to cover a large body of material. Second, each essay question on a test represents a large part of the total score. Therefore, the effect on the total score of a single unsatisfactory essay is necessarily much greater than that of one or a few short items in a large number.

The student is frequently allowed to choose one of a number of topics on which to write. If students are to be compared with one another, a choice among essay questions should not be permitted, because, unless the questions are equated for difficulty, it is difficult to judge how well the students who chose to answer questions 1, 2, and 4, for example, would have done had they chosen to answer questions 3, 5, and 6 instead. It also sometimes happens that the better students attempt the more challenging questions and write less acceptable answers than do the less able students who choose to answer the easier questions.

2.3.5 Objective Questions

Objective items have at least four advantages:

1. Since they permit the teacher to sample a greater variety of areas in a relatively short time, the student's strengths and deficiencies are more likely to reveal themselves.
2. They can be reliably scored; if the items are unambiguous and the test has been properly keyed, scoring errors will be clerical errors rather than errors of judgment.
3. They are more easily scored and, therefore, scoring time is reduced. (They can often be scored by the students themselves.)
4. They lend themselves rather readily to item analysis.[1] (Consequently, over a period of time, the teacher can assemble a file of items by keeping from each test the questions that actually discriminate well.)

2.3.6 The greatest limitation of objective questions is that they are difficult to construct if they are to test more than sheer memory. Objective questions, especially multiple-choice items, can test most of the higher intellectual abilities and skills that can be measured by essay questions. But, because the construction of objective items that do measure more than memory requires much time and considerable ingenuity, teachers are likely to be content with questions that test only knowledge of facts--sometimes very trivial facts.

---

[1] For an explanation of item analysis, see Chapter V.

2.3.7 A second limitation of objective questions is that they lend themselves to guessing on the part of the student. Guessing can be controlled by correcting for guessing, but correction for guessing is difficult for teachers to apply and there is ultimately no way to differentiate between answers that are wrong because the student has merely guessed and those that are wrong because the student has answered in good faith but on the basis of misinformation.

2.3.8 Four types of objective questions are especially common in foreign-language teaching: completion, true/false, matching, and multiple-choice. All of these types can be useful if the items are well composed. They are all easy to use in classroom tests and a large number of them can be administered in a relatively short time.

2.3.9 There is no reason why a variety of item types cannot be used in the same classroom test; the decision must depend on the purpose of the test. When more than one item type is used, however, items of the same type should be grouped so that the student is not confronted with frequent changes of directions.

2.4 LENGTH OF TESTS

2.4.1 How long a test should take depends on the purpose that the test is to serve. Except for special examinations, most longer classroom tests are limited by the length of the regular class period. (In general, however, the longer the test, the more reliable the scores.)

2.4.2 The classroom test should be deliberately planned so that most students have time to attempt all questions as they work at their normal speed. Anxiety is likely to be accentuated when there is pressure to work fast.

2.4.3 The number of objective questions that can be answered in any given length of time depends upon the questions and the work habits of the individual students; the fastest student in the class may finish a test in half the time that it takes the slowest student to finish it. For multiple-choice and completion items, many teachers find that most students can complete about 50 questions in 40 minutes; true/false and matching questions can often be answered at the rate of two per minute. It must be emphasized, however, that each objective item type is very flexible; much depends on the use the teacher makes of it. One teacher may regularly write short items of one type that students can answer quickly while another may use the same type but write items that take much longer to answer. The length of time required for essay questions depends in large measure on the complexity of the questions. The teacher can write out an acceptable answer to each essay question and multiply his time by at least three to estimate the time his students will need. If in doubt, the teacher should take the whole test himself, reading every question through completely and conscientiously writing out appropriate answers; he will then be better able to estimate the amount of working time that will be required by the students.

The student's experience of tests increases in proportion to the length of time he has spent in school. The more experience of this kind he has had, the faster he can work. The elementary-school child cannot be expected to take tests as long as those administered at the junior-high-school level; tests administered to junior-high-school students should be shorter than those administered to students in the senior high school, etc.

2.5 RELATIVE DIFFICULTY OF TESTS

2.5.1 The level of difficulty of a test should also be largely determined by the purpose that the test is to serve. An achievement test may be relatively easy; a test that is to be used primarily for grading purposes should be moderately difficult; a test that is to be used to isolate the most able students should be difficult.

2.5.2 In composing a test of minimum essentials, the teacher need not deliberately strive for a great range of difficulty; rather, he may use a large number of questions that he hopes will be answered correctly by the great majority of his students so that he can be sure that they have mastered those essentials and so that he can identify the few students who have not, for whom he must plan special remedial work. In general, such tests are not very effective in discriminating among the various levels of achievement from highest to lowest.

2.5.3 If, on the other hand, a test is to be used to identify students who may be eligible to compete for special awards or to enroll in accelerated classes, etc., the teacher will need to use more difficult questions that can be answered correctly by comparatively few students.

2.5.4 On end-of-unit tests, mid-term tests, or final examinations, it is important to be able to rank students in the order of their level of achievement. For tests of this type, many teachers like to use a few very easy items to encourage the poorer students, a few very difficult items to challenge the better students, and many items of moderate difficulty. It must be clearly understood, however, that in the last analysis it is desirable for all students to do well on all tests. If they do, there will be few C's, D's and F's, but the students will know the material they have studied. The teacher must beware lest he be carried away by considerations of grading. He must not deliberately go looking for students whom he can classify as poor! Research has shown that students react so differently to individual items that it is not necessary to vary the difficulty of items deliberately. In most cases, the more reliable and more valid tests are those in which all items are close to the 50 percent level of difficulty: i.e., each item is answered correctly by about 50 percent of the students but the variability of the students themselves in their responses to individual items is so great that a wide range of scores results.

2.6 FORMAT

2.6.1 Certain aspects of the format of a test need to be determined when the test is planned. The teacher must first decide whether the test will be presented orally or in written form. Stimulus material for tests of listening comprehension is presented orally, but it is best to present the responses in written form if they are of the multiple-choice type and each item has more than three choices. Speaking tests may have both oral and written stimuli, but the responses will be given orally, of course. Objective tests of reading and writing must obviously be presented in written form.

2.6.2 Since legibility, attractiveness, and economy of space are important, tests should be typed rather than hand-written. Responses for multiple-choice items should be listed vertically rather than horizontally whenever possible. If the answers are to be recorded directly on the pages of the test, the spaces provided for the answers to objective questions should be placed conveniently both for the student and for the teacher when he comes to score the paper. If a simple answer sheet can be designed for use with a test, the time required for scoring will be greatly reduced. It is important that items be grouped by types to keep to a minimum the number of different sets of directions. Items should also be grouped by content within each item type, and arranged from easy to difficult within the test as a whole as well as within each major subdivision of the test; such an arrangement of items makes it easier for the teacher to analyze student performance for diagnostic purposes.

2.6.3 For essay tests, wherever different amounts of credit are allowed for different questions or for different parts of a single question, the credit to be allowed should be decided when the test is planned and should be clearly indicated at appropriate places on the test. The question is frequently asked: If objective questions are specifically designed to have different degrees of difficulty, should not correspondingly different amounts of credit be given for them? For practical purposes, even though such differential weighting of objective questions might be desirable, assigning appropriately different amounts of credit is a complex process and the process of scoring such a test is complicated and time-consuming.

2.6.4 Finally, the teacher may use a variety of media as bases for test items: pictorial material shown with an opaque projector or an overhead projector, films, filmstrips, models, demonstration performances, etc. The availability of materials of these kinds can influence the choice of item types and the way in which they are presented to the students.

2.7 SCORING

2.7.1 Every test should be planned so that it can be scored as reliably as possible. For objective items scoring is easy if the items are all well constructed and if the key is correct. For essay tests scoring is more complicated; the task is simplified, however, if the questions themselves limit the student's freedom of response, if the teacher has a clear notion of the responses he will allow and has decided exactly how much credit to give for each separate part of the test.

While most classroom tests are scored by the teacher, some tests (especially shorter quizzes or tests of objective type) can be scored by the students if they are closely supervised. For objective tests, scoring is straightforward: answers are either right or wrong. For essay tests, in which subtle judgments are involved, the teacher necessarily bears the whole burden of responsibility.

2.7.2 Because essay tests are difficult to score reliably, some people consider it desirable to have more than one teacher judge a student's work. Arrangements can sometimes be made to have each student's test scored by two different teachers; the two scores can then be averaged. There is no guarantee, however, that this procedure will really reduce the ultimate subjectivity of the grade; it does make it more probable that the final score is reliable. When all teachers working at the same grade level collectively develop and give the same tests at the same time, it is possible to have each item or set of items scored by a different teacher; the final grade for each student then depends on the judgments of a number of teachers rather than on the judgment of a single teacher.

2.7.3 For the teacher who relies on his own judgment alone in grading essay questions, the following suggestions may help him achieve more reliable scores:

1. Before scoring the first paper, have a clear notion of the answers you will allow for each question and of the weight you will give to each of the various elements in the answer.
2. Score all answers to a particular question before going on to the next question.
3. Grade without knowing the identity of the student whenever possible.
4. Use categories rather than percentage grades (i.e., use

A, B, C, D, F, or 5, 4, 3, 2, 1 or Excellent, Good, Fair, Poor, Failing, etc.,--rather than 87, 76, 64, and the like).

2.7.4 The teacher can tell about the reliability of his own scoring if he grades a set of essays without recording his grades on the papers themselves and then rescores them after an interval of time and compares the two sets of results. (For this purpose, it is especially important to grade each paper without knowing whose it is.)

2.7.5 These suggestions may make the scoring of essay questions look simple. It is not. The scoring of essay questions in foreign-language tests is very complicated. Among other things, the teacher must decide how much importance to give to language and how much to content. He must also decide whether he should grade language and content separately. If he does grade them separately, how shall he combine the two grades to arrive at a single grade?

2.8 CORRECTION FOR GUESSING

2.8.1 Correction for guessing is intended to reduce to zero the chance score[2] of the student taking a test on which he makes a choice among responses.[3] A correction for guessing is applied in the scoring of many standardized tests; however, it is not really necessary to correct for guessing for most classroom tests. If every student tries to answer all the questions on a classroom test the rank order of scores will not be changed by correction for guessing.

---

[2] Chance score: the score that can be obtained by guessing alone.

[3] The formula for the correction for guessing is $R - \frac{W}{n-1} = \text{Score}$, where R represents the number of right responses, W represents the number of wrong responses, and n-1 represents 1 less than the number of choices. Thus, the formula for scoring a five-choice multiple-choice question is $R - W/4 = \text{Score}$, and the formula for scoring a true/false test is $R - W/1$ or $R - W = \text{Score}$

2.8.2 Some would argue, nevertheless, that occasional correction for guessing on objective classroom tests at least acquaints students with the effects that such correction may have on the score; the experience may be useful to them when they come to take standardized tests that are corrected for guessing.

2.9 DETERMINING A PASSING SCORE

2.9.1 We know that we cannot always write tests of the same difficulty. We know that we cannot estimate the difficulty of the questions in a test precisely enough to define a passing grade (or any other cut-off score) before the test has been administered. Therefore, the scores on any test must be scrutinized carefully and adjusted in terms of the performance of the whole group before they are interpreted or reported. Teachers must not be afraid to adjust test scores; if the scores are unexpected, the fault may lie with the teacher or the test rather than with the students.

2.9.2 It is the teacher who is responsible for measuring student achievement day by day and week by week. That measurement must be as reliable and as valid as he can make it. If a test is carefully planned, the task of writing the items and assembling them is very explicit, and reliable and valid measures do result.

Chapter III

THE CONSTRUCTION OF TEST ITEMS

3.1 ESSAY QUESTIONS

3.1.1 In composing essay questions, the teacher should be especially careful (1) to state the question itself unambiguously, and (2) to specify clearly within what limits the topic is to be treated, so that the student's efforts are clearly focused on a concrete task, and so that different students' essays on a given topic can be compared with one another.

3.1.2 Among the abilities which can be tested effectively by using carefully written essay questions are the following:

(1) to compare and contrast (people, events, objects, etc.);
(2) to develop and defend an opinion;
(3) to discern and explain a cause or an effect;
(4) to summarize;
(5) to analyze complex phenomena into their component parts;
(6) to give examples to illustrate rules, principles;
(7) to criticize the adequacy, relevance, accuracy, of a given statement, assertion, opinion;
(8) to reason inductively or deductively.

3.1.3 The beginning or intermediate student can be asked to write more or less lengthy and complex pieces of prose according to his ability; he can be asked, for example, simply to state clearly and straightforwardly certain information he is presumed to possess, or to describe accurately a person, object or place with which he is familiar, etc.

3.2 OBJECTIVE ITEMS

3.2.1 Multiple-Choice Items

3.2.1.1 A multiple-choice item may take several forms: (1) it may be a direct question followed by several possible answers; (2) it may be an incomplete sentence (called the *stem*) followed by several possible completions; or (3) it may be a problem, a graph, a diagram, etc., followed by several statements. The student may be asked to select the one choice that is *correct*, the one that is *incorrect*, or the one that is *best*. These three kinds of answers combined with the three item forms give the test constructor nine possibilities. Beginners are usually more successful at writing *questions* than at composing items consisting of stem and completions.

Multiple-choice items can be used to measure the student's ability to discriminate, interpret, analyze, make inferences, and solve problems. It is sometimes argued that multiple-choice items are inherently weak in that students are required merely to recognize and judge proposed solutions, interpretations, or distinctions; such items are weaker, it is contended, than others in which the student is required first to recall possible answers and then to select the correct one. Others argue that the kinds of problems posed by multiple-choice items come much nearer to real life than do those of any other type of test question. They contend that people are rarely called upon to provide all possible solutions to a problem and then choose among them; rather, several alternative possible solutions are frequently apparent and the essential problem is to identify the best or correct solution. Studies have shown that the ability to recall correct solutions correlates highly with the ability to recognize correct solutions; it seems justifiable, therefore to continue to make widespread use of carefully constructed multiple-choice items.

Critics of multiple-choice items often base their criticism on the assumption that such items can test only facts or definitions. In practice, such criticism is too often justified: the teacher who uses multiple-choice test items frequently finds it easier to write only items of this type. It takes more time, imagination, and ingenuity to develop test items that measure a student's ability to interpret, to draw inferences, to apply knowledge, or to think critically. Multiple-choice items can be written, however, to test these abilities with great subtlety. Carefully written multiple-choice items provide high validity and reliability and they are easy to score.

3.2.2 Construction of Multiple-Choice Items

3.2.2.1 Three general difficulties are encountered in the construction of multiple-choice items: (1) the language of the item must be clear and unambiguous; (2) the answer must be unquestionably correct; and (3) the distractors must be attractive to those students who do not possess the knowledge or understanding necessary to recognize the correct answer.

3.2.2.2 Here is a list of specific points that the teacher should bear in mind as he writes multiple-choice items:

(1) The directions should tell the student explicitly whether more than one answer is possible and whether he is to select the correct answer or the best answer.

(2) The stem or question should be worded simply and understandably. Use words appropriate to the subject matter and to the group being tested. Be as succinct as possible. The solution to the question should not depend upon the student's ability to understand unnecessarily complicated language.

(3) Each item should pose one question; do not test two or more points in a single item.

(4) Each item should be independent of all other items in the same test. Students should not be able to select the correct answer to one question because of information gleaned from another item, and they should not be penalized several times for missing one item: if a student misses one item he will also miss all others that depend on it.

(5) The choices should be as short as possible. Words repeated in each of the alternatives can usually be added to the stem; a question can often be reworded to simplify the responses.

(6) The best distractors are based on common mistakes made by the students. Excellent distractors can often be derived from wrong answers to essay questions and completion items.

(7) Use four or five choices whenever possible. This reduces the chance that the student will guess the correct answer.

Use the same format throughout a test.

(8) All distractors should seem equally plausible to the student who does not know the correct answer or cannot arrive at it logically. (If, for example, two of four alternatives are obviously wrong, the question becomes a true/false item: one of the remaining alternatives is the correct answer.)

(9) It is best to use random order for the positions of the correct answers. Students are quick to perceive patterns or apparent patterns.

(10) Do not make the correct response consistently longer or shorter than the distractors. (It often happens that the teacher makes the correct answer more detailed than the others.)

(11) If plausible distractors are difficult to find, use another type of item.

(12) In using stems that are incomplete sentences, write the item so that the missing part of the sentence is to be added at the end rather than at the beginning or within the sentence.

3.2.3 True/False Items

Many teachers believe that true/false items are the best type of item to measure easily and accurately the student's knowledge of specific facts.

3.2.4 Construction of True/False Items

3.2.4.1 Here are some important considerations that the teacher should bear in mind when he prepares true/false items:

(1) If the answers are to be indicated on the test paper itself, have the student *circle* the proper response. (A common practice is to require the student to write T or F, t or f, + or -, or + or 0, for true and false, respectively. It is often very difficult to distinguish between such marks, especially when erasures are permitted.)

(2) State each item clearly and specifically.

(3) A true/false item should deal with a single definite topic. Whether the item is to be judged true or false should depend upon an *important* aspect of that question.

(4) In a two-part item, the crucial element should come at the end. The first part should set the problem. (For example, in an item designed to evaluate students' understanding of the effect of a given cause, put the effect first.)

(5) Use *approximately* equal numbers of true items and false items.

(6) Avoid words that give irrelevant clues to the answer. Such words (called *specific determiners*) enable the student to answer correctly without possessing the specific knowledge in question. (Studies have shown that in the great majority of cases, sentences containing such words as *always*, *no*, *never*, *all*, *none*, etc., are most often false, while statements with such moderate words as *some*, *may*, *often*, *generally*, etc., are most often true. If you are careful to balance the number of true and false statements containing terms of these kinds, the "determining" effect will be greatly reduced. Wherever possible, it is best to avoid such words altogether.)

(7) Use quantitative rather than qualitative language wherever possible. (Terms such as large, many, important, better, etc., should be avoided. Such expressions are ambiguous and have very different values for different individuals.)

(8) In typing up the final copy of a test double-space true/false items. (If true/false questions are written too close together they can be hard to read and confusing to answer.)

(9) Avoid compound statements consisting of two or more essentially independent parts. (If one part is true and another false, the statement is neither wholly true nor wholly false. The student cannot mark such an item true or false unless he is given specific directions for marking items which are only partially true. Often in such items each part really deals with a separate problem. In such cases the question should be broken up into two items, each dealing with a single fact or idea.)

(10) Avoid double negatives. (Students who know the information involved in a question may be confused unnecessarily by double negatives.)

(11) Do not use word-for-word statements extracted from textbooks, syllabi, lecture notes, etc. (Often, when such statements are taken out of context, they are ambiguous. A true/false item that presents a significant fact or a generalization in a new context is less likely to test mere recall.)

3.2.5 Modified True/False Items

A. The effect of guessing on true/false items can be reduced by increasing the number of possible answers to the question. Instead of asking the student to choose between true or false, he can be asked to choose among three possible answers: true, uncertain, or false; correct, partially correct, or incorrect; agree, undecided, or disagree; etc. Such a device reduces to one in three the chance that the student will guess the correct answer and it provides more information than the standard true/false form.

B. The corrected true/false item is a modification designed to reduce guessing and to direct the attention of the student to the crucial element in the statement. The crucial element of the statement is underlined. The student is directed to pay attention to this key word or phrase and to use it as the basis on which to decide whether the statement is true or false. If he decides that the statement is false, he must then correct it by substituting an appropriate word or phrase for the underlined crucial element so that the statement is true. This form implies two possible scores for each item: one point for a correct answer without appropriate correction and two points for a correct answer with appropriate correction. Since items of this type take more time than simple true/false items, fewer of them can be used in a given time.

C. A third modification also reduces guessing and increases information, but at the expense of increased difficulty in grading: the standard true/false form is used, but the student is also required to state, in a few short sentences, why the statement is true or false. Scoring can be adjusted to cover all possibilities.

3.2.6 Matching Sets

3.2.6.1 Matching sets provide a convenient way to measure knowledge of series of facts, principles, relationships, or interpretations. Matching sets usually consist of material arranged in two columns: the items in the first column provide stimuli and the items in the second column serve as responses. The student's task is to select the one response which is most closely related to each stimulus.

3.2.6.2 Matching sets may be thought of as condensed series of multiple-choice items. A matching set with six pairs of related terms can be used to measure the same thing as six separate but related multiple-choice items. These six items would require many more words and more time for the student to read and answer the questions.

3.2.6.3 The principal advantage of matching sets, therefore, is their ability to measure quickly the student's ability to discriminate among several similar items as they are related in a given way. Matching sets are relatively easy to write, but they are often very poorly conceived. Nevertheless, ease of construction is also one of the major advantages of matching sets. Other advantages are their applicability to many different kinds of subject matter, their comparative freedom from guessing effect, and their minimal dependence on the student's reading ability or reading speed.

3.2.6.4 Major disadvantages of matching sets are that it is easy to fall into the habit of composing items of this type that overemphasize mere recall, and that it is easy to give inadvertent clues to the correct answers.

3.2.7 Construction of Matching Sets

Many of the precautions mentioned in the discussion of the construction of multiple-choice and true/false test items are equally applicable here (use simple, clear statements, be grammatically consistent, define the problem clearly, avoid giving clues, etc.). The following additional considerations should be borne in mind in writing matching sets:

(1) The stimulus column should be on the left, the response column on the right. Each of the terms in the left-hand column should have a number (each of them is a separate item). The items in the right-hand column should be lettered. When no separate answer sheet is used, have the student place his answer to each item in a space to the left of the number of the item. The use of a number for each stimulus and a letter for each response simplifies directions.

(2) It is important that the items listed in each of the columns be homogeneous. If the terms listed in the response column are not homogeneous, the student may be provided with clues which will help him to match the terms in the two columns. The items are then proportionately easier. The selection of the correct term should depend on the student's knowledge of the relationship being tested rather than upon his ability simply to eliminate incorrect answers. (If the items are heterogeneous, there is also the possibility that the student may discern a relationship other than the

one the examiner intends; he can then argue that his answers are also correct.)

(3) In writing matching sets in which the student is to match single words with long definitions, names of persons with quotations, or other single words or short phrases with long phrases, put the long phrases (definitions, quotations, etc.) in the stimulus column and the single words or short phrases (names, dates, etc.) in the response column. This arrangement reduces the amount of time required to answer the question; the student rereads the shorter rather than the longer list of elements to make his choice.

(In matching sets involving terms and definitions, the student is usually given the definition and he must select from among a number of possibilities the term it defines. Many teachers prefer to give the student the term and ask him to choose a definition; for this purpose the order suggested above must obviously be reversed: the short element--the term--appears in the left-hand column and the longer elements--the definitions--in the right-hand column.)

(4) To reduce the effect of guessing, the response column should contain a few more elements than the stimulus column. The extra elements should be homogeneous with the other response elements, if they are to serve as effective distractors.

(5) Experience has shown that the stimulus column should not contain more than ten elements. In instances where many more homogeneous stimulus and response items are possible and desirable on a test, it is best to divide them into a number of matching sets. By decreasing the reading time this reduces the time needed to locate

the correct answer.

(6) To enable the student to work efficiently, the items in the response column should be placed in some logical order. For example, dates should be listed chronologically; names or places should be listed alphabetically, etc. A logical order enables the student to scan the list quickly to see whether the name or term he has in mind is among those listed. A random order of elements merely increases reading difficulty unnecessarily.

(7) All parts of a matching set--the directions and the stimulus and response columns--should be on the same page. The student should not have to turn the page to match stimulus items with response items. This increases the difficulty of taking the test and may actually decrease the efficiency of the test item for measuring purposes.

3.2.8 Because it is difficult to write matching sets to measure more complex mental abilities, most matching sets in practice test knowledge of correlations between events and dates, events and places, individuals and events, individuals and quotations, or other specific facts. More imaginative uses of matching sets require the student to match causes and effects, principles and applications, situations and judgments, or problems and solutions. The student can also be asked to match places or events and locations on a map, descriptions or names and parts of a piece of equipment shown in a picture, etc.

Chapter IV--FRENCH

PREPARING TEST ITEMS

4.1 INTRODUCTION

4.1.1 It is not easy to amass a large repertoire of effective test items. In addition to the difficult task of writing understandable instructions and credible distractors, the teacher must worry constantly about how the test fits what he has taught. In foreign-language work, it is often desirable for a test to sample all the language skills; testing devices must be used which are appropriate to each skill. We hope that the teacher will find the following sections helpful in making his own test items. This manual claims neither novelty nor completeness. We have tried, first, to be explicit about the essential elements of successful foreign-language tests and then to organize them into a useful working outline accompanied by concrete examples.

4.2 THE ESSENTIAL ELEMENTS OF A TEST

4.2.1 There are three basic sets of elements in any language test: (1) the language skills, (2) specific testing devices, and (3) the particular language areas to be tested.

4.2.2 The Language Skills

4.2.2.1 Teachers generally recognize four fundamental skills: (1) listening comprehension, (2) speaking, (3) reading, and (4) writing. In audio-lingual programs, listening comprehension is considered to have primacy over the others. The position taken is that one cannot speak, read or write meaningfully in a language without being able to understand that language. The other three skills are also interrelated, but it is not clear that speaking is as important for reading and writing as listening comprehension is for all of them. Even in the native language, the reading skill is often--perhaps always--more highly developed than the ability to write.

4.2.2.2 Tests must take into consideration which specific skills are to be evaluated. When the reading skill has been the main concern of the course, for instance, tests should not emphasize listening comprehension; in audio-lingual programs, which generally begin with listening comprehension and speaking, reading and writing should not be tested in the early stages of instruction.

4.2.2.3 Of the four skills, listening comprehension and reading are easiest to evaluate objectively. The difficulties inherent in testing speaking and writing are readily apparent; with listening and reading it is a comparatively simple matter to control the student's responses, but it is difficult to do so when the student is to react orally or in writing. Judgment of spoken and written responses involves much personal interpretation by the teacher. The burden of grading and administering oral-response tests makes them unwieldy; although written answers are somewhat easier to grade, they are just as difficult to control as spoken responses. Having a number of teachers grade written responses may actually increase the degree of subjectivity, although some experts disagree.[1]

4.2.3 Testing Devices

4.2.3.1 The most commonly used devices in foreign-language testing are: (1) multiple-choice items, (2) completion items, (3) matching sets, (4) true/false items, (5) translation, (6) dictation, and items involving (7) expansion, (8) transformation, and (9) substitution. The last three are especially useful to test speaking and writing. Of the other devices (see Chapter III), true/false items and translation are the ones most often debated. It is very difficult to phrase a true/false item so that it is clearly and unquestionably true or false but not transparently so. Translation has several shortcomings. Using translation from the native language into the foreign language to discover the student's control of specific problem areas (the subjunctive, for example, in French) is ineffective in cases where the student can find an alternate possibility (the infinitive, for example). Although translation from the foreign language to the native language is easier to control, its effectiveness is also somewhat limited because the student can usually find clues to the meaning of the phrase or word to be translated in language features (context, word order, etc.) other than the specific feature being tested. Therefore, unless the teacher can control quite strictly the context of a particular word or phrase, it is difficult to determine the student's actual command of the matter being tested. Extensive use of translation is obviously subject to attack in any curriculum that consciously emphasizes direct use and knowledge of the foreign language with as little recourse as possible to the student's native language. Perhaps translation is more accurately a specialized skill than a testing device. There is certainly no proof that the ability to understand, speak, read, and write a foreign language also guarantees the ability to translate freely between that language and one's own language. Indeed, in institutions with a sound foreign-language program designed for specialists, translation is often taught in a separate course to which students are admitted only after they have demonstrated a command of the other skills.

---

[1] See, for example, The Measurement of Writing Ability by Fred Godshalk, Frances Swineford and William Coffman, College Entrance Examination Board, New York (1966). The articulation of norms for written composition differs from language to language. For an illuminating discussion of procedures, see How the French Boy Learns to Write by Rollo Walter Brown, Harvard University Press (1915), reprinted by the National Council of Teachers of English (1963).

4.2.3.2 Multiple-choice items (i.e., choice of one correct answer from two or more possible answers) and matching sets are easy to control since variation in answers is strictly limited. To be effective, multiple-choice items must have distractors which are sufficiently similar to the correct choice to attract the student who does not really know what the correct answer is. Incorrect answers can reveal the individual student's problems. For example, a wrong choice, on an item testing listening comprehension, between le ton and le temps, l'amant and la main may be symptomatic of a general inability to distinguish nasal vowel phonemes from one another. By increasing the number of possible answers, the teacher can gain more information about the individual student's weaknesses and needs.

4.2.3.4 There are two ways of constructing *matching sets*: (1) the same number of items may be supplied in each of two columns, or (2) more items may appear in the column from which the 'answers' are to be selected, thereby leaving a residue when the answers have been chosen. In either case, matching is a type of interlocking multiple choice. Guessing, however, can be a factor in both matching and multiple-choice items. A formula for penalizing random choices can be used; it should be explained to the students in an effort to discourage them from making irresponsible guesses.

4.2.3.5 *Expansion items* test the student's knowledge of the construction of sentences in the foreign language. The student may be asked to expand a sentence in a number of different ways. In the following example, he is being tested on his knowledge of the positioning of adjectival modifiers in noun phrases:

Instructions: In each of the following sentences insert the word you are given:

Sentence No. 1: Les femmes étaient arrivées.

Insert: jolies

Answer: Les jolies femmes étaient arrivées.

Sentence No. 2: Les dames causaient.

Insert: distinguées.

Answer: Les dames distinguées causaient.

4.2.3.6 Transformation items are particularly suited to evaluating writing. The student is given instructions and a model; for example, he is given a sentence with a verb in the present tense to be changed to a past tense. A series of sentences follows with the verb in the present tense which the student is to change to the past. Transformation items can be written to test more complex matters such as making active sentences passive, inserting subordinate clauses into simple sentences to make complex sentences, etc. It is always advisable to furnish a model item as part of the instructions.

4.2.3.7 In substitution items, the student is asked to replace a word, part of a word, or phrase with another:

Instructions:	Replace the underlined words with others which are equally correct.
Sentence:	Je me suis lavé les mains.
Possible Answer:	Je me suis lavé la figure.

4.2.3.8 Dictation is a well-known device in foreign-language instruction. However, it is difficult to grade straight dictation. The teacher must decide how to sort out and weight each student's individual errors. One way to avoid many problems is to give the student a partially filled-in answer sheet. As the teacher reads the dictation, the student writes in only what has been omitted from the text he has before him. The blanks may become more numerous as the student progresses in his ability to take dictation. Dictation has traditionally played an important rôle in the teaching of French. It is most useful as a testing device, to test the student's ability to understand oral French and his knowledge of French orthography.

4.2.4 The Language Areas

4.2.4.1 The third basic set of elements involved in the preparation of foreign-language tests, the language areas to be tested, is the least clearly defined and by far the largest of the three. Opinions about the relative importance of various language areas differ widely.

4.2.5 Contrastive Analysis

4.2.5.1 The student's native language is carefully contrasted with the foreign language he is studying because it is believed that the language learner will make numerous predictable errors when he incorrectly transfers the habits he has acquired in his native language to the language he is studying. In other words, the learner tends to express himself in the foreign language in terms of the rules of the language he already knows. He will automatically assume that whatever is not clearly different in the foreign language is the same as in his own language. A thorough knowledge of the differences that exist between the two languages can therefore help the teacher to anticipate problems.

4.2.5.2 A foreign-language test could well consist very largely of items on the areas where the two languages differ. If the teacher knows, for example, that French grammatical gender is different from English gender, he can expect gender to be a problem for the English-speaking learner and may therefore devote a large proportion of his testing to gender in the early stages of instruction.

4.2.5.3 Wherever the native language of the student has a feature which is lacking or simpler in the foreign language, the problem is relatively easy to solve. Where French has n'est-ce pas, for example, English has a far more complicated set of expressions. As a result, the English-speaking student of French has to learn one invariable expression while the French-speaking student of English has to learn to use correctly the set of English expressions aren't we, couldn't we, etc.

4.2.5.4 Both the contrasts and the ways in which we state them depend on the theoretical premises we employ in our contrastive analysis. We suggest the following outline as a good guide to the arrangement of basic language matters:

- I. Phonology (Sound System)
    - a. phonemes (distinctive classes of sounds)
    - b. distribution of phonemes (including diphthongs and consonant clusters)
    - c. intonation (pitch changes in the voice to indicate statements, questions and emotional coloring)
- II. Morphology (Formation of Words)
    - a. morphemes (the basic units of grammar)
    - b. inflection (changes in words according to number, person, case, etc.)
    - c. derivation of words (formation of new words by compounding or by the addition of prefixes and suffixes)
- III. Syntax (Arrangements of Words in Larger Constructions)
    - a. phrases
    - b. clauses
    - c. sentences

4.2.6 Vocabulary

The size and nature of a student's vocabulary is obviously important in language teaching. Various word counts exist,[2] but there is surprisingly little agreement about what words should be taught at each level of language instruction. However, since the teacher must test vocabulary, he should select vocabulary items as shrewdly as he can to test those items that seem most useful. The studies carried out in France to select vocabulary for *Le Français fondamental*, and related courses, have yielded sophisticated vocabulary lists that the teacher will find useful in making up vocabulary tests. For purposes of classification, idioms should be included in 'vocabulary'. It is also often useful to group words into families or general categories.

4.2.7 Style

Speakers of French make systematic use of a number of different styles of discourse of varying degrees of formality and impersonality, both orally and in writing. Some distinctions in style affect even basic grammatical forms. Such is the case with the several past tenses, both indicative and subjunctive, choice among which is governed to some extent by considerations that are properly *stylistic*. At least the most important distinctions of style are treated in most language texts and should be tested.

---

[2] An interesting study has been done for English at Brown University (see the forthcoming publication, *Computational Analyses of Present Day American English*, edited by Henry Kucera and W. N. Francis). In a text of 1,014,235 words, there were only 50,457 different words; 42,540 of these occurred only ten times or less.

4.2.8 Literature

Literature is generally considered an essential part of foreign-language instruction. In audio-lingual programs literature is not ignored; rather, it is not treated until students can read with some ease and with sensitivity to literary values. The choice of specific literary selections usually varies, of course, from teacher to teacher and from class to class, although in some instances detailed syllabi specify the texts to be read in a given course.

4.2.9 Culture

There is no generally accepted, useful definition of culture in the context of foreign-language teaching. Culture is understood either in a narrow sense, meaning understanding of the literary and artistic values general among the educated speakers of a language, or it is extended to include the whole way of life and the history of a people, the geography of the land they live in and their political, religious and social institutions. Teachers seem to agree, however, that culture is an essential part of language instruction and that they should test their students on it.

4.2.10 A Brief List of Language Areas

The following list of language areas is offered:

(1) distinctive sound classes (phonology) (e.g., the vowel sounds in *la*, *les*, *lit*, *lot*, *lu*);

(2) diphthongs and combinations of vowels and /w/ or /j/ (e.g., *puis*, *oui*, *oie*, *travail*);

(3) consonant clusters (e.g., consonant + /l/ or /r/ in final position: *pauvre*, *titre*, *aimable*, *mufle*);

(4) statement intonation (e.g., the even pitch level in French);

(5) Question intonation (e.g., *Va-t-il à Paris*? vs. *Qui va à Paris*?);

(6) basic grammar units (e.g., *à* vs. *de*, *avoir* and *être* as auxiliaries);

(7) inflectional procedures (e.g., verbal inflection);

(8) derivation (e.g., nouns of agent derived from verbs);

(9) phrasal syntax (e.g., the positioning of adjectives: *un pauvre homme*, *un homme pauvre*);

(10) clausal syntax (e.g., use of subordinate clauses introduced by *que*, etc.);

(11) sentence syntax (e.g., sequence of tenses);

(12) vocabulary (including both single words and idioms like *faim*, *avoir faim*);

(13) style (e.g., use of certain past tenses in formal styles);

(14) literature (names of well-known writers, reading selections);

(15) culture (accepted ways of speaking in specified social situations, ways of addressing officials, etc.).

The list is open; additional features can be added to it and the ones given can be made more detailed.

4.3 WORKING OUTLINES FOR TEST WRITING

The skills and various testing devices have been arranged on the following chart:

Chart No. 1: Skills and Testing Devices

Skills	Testing Devices									
	MC	CO	MA	T/F	TR	DIC	EX	T	SUB	RA
Listening Comprehension	X		X	X	X					
Speaking	X	X			X		X	X	X	X
Reading	X	X	X	X	X			X		X
Writing	X	X			X	X	X	X	X	

Key:
MC - multiple choice
CO - completion
MA - matching
T/F - true/false
TR - translation
DIC - dictation
EX - expansion
T - transformation
SUB - substitution
RA - reading aloud
X - effective use

An examination of the chart leads to some general observations on the appropriateness of particular testing devices to specific skills:

(1) Generally speaking, completion items are hard to control. In order to assure objectivity, completion items must be restricted to absolutely unambiguous possibilities. (Visuals can be useful for this purpose.)

(2) Ambiguity is a major problem in true/false questions. The teacher should be especially cautious in his use of ungrammatical sentences which the student is to identify as "incorrect." Just as correct stimuli appear to reinforce learning, incorrect stimuli may

2

create uncertainty where the student is well prepared.

(3) The use of translation to test all four skills is open to question. Problems arise in restricting translation so as to test only specific points. Translation can also be considered a distinct skill; as such, it, too, can be tested by a number of devices, such as multiple choice, completion, matching and true/false items.

(4) Expansion and transformation items are useful devices for testing speaking and writing, but they are inappropriate devices for testing the receptive skills of listening and reading.

It is reasonable to suppose that some testing devices are more suitable than others to pupils of different ages. In the absence of evidence from controlled experimentation, the individual teacher must rely on his own judgment.

4.3.1 Chart No. 2, combining language areas and skills, is a check list for the preparation of test items. By filling it with actual test items of different kinds, the teacher can achieve a broad coverage of language problems. The following samples illustrate item composition. Each item is numbered in cross reference to Chart No. 2.

Chart No. 2: Language Areas and Skills

Language Areas	Skills			
	Listening Comprehension	Speaking	Reading	Writing
Sound Classes (Phonemes)	(1), (2)			
Diphthongs (and longer chains of vowels)				
Consonant Clusters				
Statement Intonation	(3), (4)	(4)		
Question Intonation	(3), (4)	(4)		
Grammar Units (Morphemes)	(7)		(5)	(5)
Inflection		(6)		
Derivation			(8)	(8)
Syntax: phrases		(10)	(10)	(9)
Syntax: clauses		(10)	(10)	
Syntax: sentences		(11)		
Vocabulary	(13)	(17)	(12)	
Style	(14)			
Literature				(15)
Culture	(16)			

4.3.2 Sample Test Items

(1) **Language Area:** Sound discrimination (Phonology)

Skill: Listening comprehension

Testing Device: Multiple choice

Instructions: Listen to the following sets of three words each. One word of the three is different from the other two. Indicate by circling on your answer sheet whether it is *a*, *b*, or *c* which is different from the other two.

Student hears:

1. a. bon
   b. beau
   c. bon

2. a. pan
   b. pan
   c. pas

3. a. plein
   b. plein
   c. plane

Student sees on answer sheet:

1. a b c
2. a b c
3. a b c

Answers:

1. b
2. c
3. c

Comment:

Sets of words of this kind can be adapted easily to test the student's ability to make phonologically essential distinctions in speech.

(2) Language Area: Sound discrimination

Skill: Listening comprehension

Testing Device: True/False

Instructions: You will hear several pairs of words or expressions. Indicate on your answer sheet by circling s or d to indicate whether these pairs are the same or different.

Student hears:

1. Je me lève. Je me lave.
2. Il s'appelle. Il s'appelle.
3. C'est un bon chien. C'est un beau chien.

Student sees on answer sheet:

1. s d
2. s d
3. s d

Answers:

1. d
2. s
3. d

Comment:

In this context s means 'It is true that they are the same' and d means 'It is false that they are the same; they are different.'

(3) Language Area: Intonation

Skill: Listening comprehension

Testing Device: True/False

Instructions: Look at the following expressions. Add a question mark or a period to each one as you hear it.

Teacher reads: (The pupil's text is not punctuated.)

1. Moi aussi (falling intonation)
2. Tu y vas (rising intonation)

3. Que fait-il (falling intonation)
4. Moi aussi (rising intonation)
5. Tu y vas (falling intonation)

Answers:

1. Moi aussi.
2. Tu y vas?
3. Que fait-il?
4. Moi aussi?
5. Tu y vas.

Comment:

This item tests intonation and involves listening comprehension skill. The student is being asked to make a simple binary decision: if the sentence is interrogative ("True") he adds a question mark; if it is not ("False"), he adds a period.

(4) Language Area: Intonation

Skill: Speaking

Testing Device: Transformation

Instructions: Make a question out of each of the following statements without adding or changing any words.

Student hears:

1. Vous allez au cinéma. (falling intonation)
2. Il part. (falling intonation)
3. Le repas est terminé. (falling intonation)

Answers:

1. Vous allez au cinéma? (rising intonation)
2. Il part? (rising intonation)
3. Le repas est terminé? (rising intonation)

Comment:

This item involves listening comprehension as well as the ability to use intonation patterns appropriately.

(5) Language Area: Basic grammar

Skill: Writing

Testing Device: Transformation

Instructions: Rewrite the following sentences, using pronouns for the underlined words and making any other changes necessary.

Example: Je donne le livre au garçon.
Je le lui donne.

1. Il met le cahier dans le tiroir.
2. Donnez du lait au chat.
3. Nous parlons de ces livres à nos amis.
4. Elle envoie la lampe à ses parents.
5. Je vais expliquer la leçon.
6. Ne dites pas la vérité à Jeanne.

Answers:

1. Il l'y met.
2. Donnez-lui-en.
3. Nous leur en parlons.
4. Elle la leur envoie.
5. Je vais l'expliquer.
6. Ne la lui dites pas.

Comment:

This item tests grammar (use and position of object pronouns). It involves reading. It could be administered orally to test listening comprehension and speaking. In scoring this item, one point might be given for each correct pronoun and a second point if it is correctly placed.

(6) Language Area: Verb inflection

Skill: Speaking

Testing Device: Transformation

Instructions: Listen to the following verbs, then change them from the plural to the singular.

Student hears:

1. ils partent
2. ils rompent
3. elles descendent
4. elles dorment

Answers:

1. il part
2. il rompt
3. elle descend
4. elle dort

Comment:

This item tests grammar (verb inflection). This item could be reversed; it would then be more difficult because the pupil would have to supply the correct final consonants.

(7) Language Area: Basic grammar (tenses)

Skill: Listening comprehension

Testing Device: Multiple choice

Instructions: Circle one of the following words: hier, aujourd'hui, or demain, to correspond to the time of each of the statements you will hear.

Student hears:

1. J'ai mangé une pomme.
2. Il est malade.
3. Nous arriverons à dix heures. (or: Nous allons arriver à dix heures.)

Student sees on answer sheet:

1. hier aujourd'hui demain
2. hier aujourd'hui demain
3. hier aujourd'hui demain

Answers:

1. hier
2. aujourd'hui
3. demain

Comment:

This item incidentally involves reading.

(8) Language Area: Derivation

Skills: Writing/Reading

Testing Device: Completion

Instructions: Fill in the blanks to complete the following sentences:

Example: L'épicier et l'______________ travaillent dans une épicerie.

L'épicier et l'epicière travaillent dans une épicerie.

1. Le ______________ et la ______________ travaillent dans une blanchisserie.

2. Le ______________ et la boulangère travaillent dans une ______________.

Answers:

1. blanchisseur
   blanchisseuse

2. boulanger
   boulangerie

Comment:

This item may be administered orally to test listening comprehension and speaking.

(9) Language Area: Phrasal syntax

Skill: Writing

Testing Device: Transformation

Instructions: Change all nouns to the feminine, making any other necessary changes.

1. Ce bon acteur est mon ancien élève.
2. Les japonais sont très sportifs.
3. Son fiancé est couturier.

Answers:

1. Cette bonne actrice est mon ancienne élève.
2. Les japonaises sont très sportives.
3. Sa fiancée est couturière.

Comment:

This item might be scored by giving one point for each correct noun change, one point for each correct adjective change, and one point for each correct change of any other kind.

(10) Language Areas: Clausal syntax/Phrasal syntax

Skills: Speaking/Reading

Testing Device: Reading aloud

Instructions: Read the following sentences with normal intonation.

1. NOUS AVONS DEUX AMIS.
2. IL Y A TROIS HEURES QU'ILS ATTENDENT.

Answers:

1. Nous‿avons; deux‿amis.
2. Trois‿heures; ils‿attendent.

Comment:

This item tests the student's ability to recognize phrases and clauses as units and to make liaisons as required by the internal structure of such combinations.

(11) Language Area: Sentence syntax

Skill: Speaking

Testing Device: Transformation

Instructions: Repeat each of the following sentences; give its corresponding interrogative form and negative form.

Student hears:

1. Tu as un crayon.
2. Il mange des frites.
3. Nous prenons une glace.

Answers:

1. As-tu un crayon?
   Tu n'as pas de crayon.

2. Mange-t-il des frites?
   Il ne mange pas de frites.

3. Prenons-nous une glace?
   Nous ne prenons pas de glace.

Comment:

This item incidentally involves reading.

(12) Language Area: Vocabulary

Skill: Reading

Testing Device: Multiple choice

Instructions: In each of the following items, circle the letter of the expression which would least change the meaning of the sentence if substituted for the part of the sentence that is underlined.

1. Je regrette, Madame, mais Monsieur le directeur est actuellement en vacances.

   a. en ce moment
   b. véritablement
   c. enfin
   d. assurément

2. J'ai bien aimé la conférence d'hier soir.

   a. la lecture
   b. le discours
   c. la communication
   d. la réunion

Answers:

1. a
2. b

(13) Language Area: Vocabulary

Skill: Listening comprehension

Testing Device: Multiple choice

Instructions: Write on your answer sheet the letter of the expression which most reasonably completes each of the statements you hear.

Student hears:

1. Dans la cave on ne voit pas bien.

    a. Elle est vieille.
    b. Elle est grande.
    c. Elle est sombre

2. En France, la distance est mesurée . . .

    a. en mètres.
    b. en kilomètres.
    c. en miles.

3. Cet homme travaille toute la journée dans un atelier d'usine.

    a. C'est un ouvrier.
    b. C'est un paysan.
    c. C'est un sculpteur.

4. Cet homme travaille dans un magasin.

    a. C'est un professeur.
    b. C'est un boulanger.
    c. C'est un employé.

5. Cet homme travaille dans le bloc opératoire d'un hôpital.

    a. C'est un chirurgien.
    b. C'est un ingénieur.
    c. C'est un cordonnier.

Student sees:

1. a b c
2. a b c
3. a b c
4. a b c
5. a b c

Answers:

1. c
2. b
3. a
4. c
5. a

(14) Language Area: Style

Skill: Listening comprehension

Testing Device: Multiple choice

Instructions: You will hear sets of three utterances each. Circle the letter corresponding to the most polite utterance of the three.

Student hears:

1. a. Pardon, Monsieur. Pouvez-vous me dire à quelle heure ouvre le magasin?

   b. Pardon, Monsieur. Pourriez-vous me dire à quelle heure ouvre le magasin?

   c. Pardon, Monsieur. A quelle heure ouvre le magasin?

2. a. Asseyez-vous, s'il vous plaît.

   b. Veuillez vous asseoir, s'il vous plaît.

   c. Voulez-vous vous asseoir, s'il vous plaît?

3. a. Voudriez-vous m'expliquer ce problème?

   b. Voulez-vous m'expliquer ce problème?

   c. Expliquez-moi ce problème.

Answers:

1. b
2. b
3. a

Comment:

This item tests style; it tests the student's ability to recognize a proper fit between language and social situation.

(15) Language Area: Literature

Skill: Writing

Testing Device: Completion

Instructions: Complete each of the following sentences.

1. Dans le Tartuffe, Molière satirise ______________.
2. Dans le Bourgeois Gentilhomme, Molière satirise ______________.
3. Dans les Précieuses Ridicules, Molière satirise ______________.

Possible Answers:

1. l'hypocrite réligieux
2. le nouveau riche
3. les fausses savantes; la préciosité

Comment:

This item tests knowledge of specific facts of literary works. Note that the student has a certain latitude in filling in the blanks. The teacher must accept any answer that is both factually and linguistically correct; he may decide to give part credit for correct information but faulty French or for acceptable French expressing inaccurate information.

(16) Language Area: Culture

Skill: Listening comprehension

Testing Device: Multiple choice

Instructions: Circle on your answer sheet the letter corresponding to the expression that most appropriately completes each sentence.

Student hears:

1. La France est bornée au sud par ___________

   a. l'océan Atlantique.
   b. la mer Méditerranée.
   c. la Manche.
   d. le Finisterre.

2. La Seine se jette dans ___________

    a. l'Atlantique.
    b. le Pas de Calais.
    c. la Méditerranée.
    d. la Manche.

3. Bordeaux est situé sur ______________

    a. la Garonne.
    b. la Loire.
    c. le Rhône.
    d. la Saône.

4. Les montagnes qui forment une partie du massif central sont ______________

    a. les Alpes.
    b. les Pyrenées.
    c. les Vosges.
    d. les Cévennes.

Answers:

1. b
2. d
3. a
4. d

Comment:

These items test culture in the broad sense. The student has only numbered series of letters on his answer sheet (1. a, b, c, d) and the teacher reads (one time only) each stem and the four choices corresponding to it. For any listening test, the teacher should write on the blackboard the large capital letters A B C D and point to each letter as he reads the corresponding choice. Items of this type can also be used in written form to test reading.

(17) Language Area: Vocabulary

Skill: Speaking

Testing Device: Transformation and completion

Instructions: Answer each question in French.

Student hears:

1. Dans une boucherie, qu'est-ce qu'on achète?
2. Dans une boulangerie, qu'est-ce qu'on achète?
3. Dans une charcuterie, qu'est-ce qu'on achète?

Comment:

Items of this type involve listening comprehension. They test knowledge of specific vocabulary and the student's ability to transform part of the question into part of his answer. Since the teacher will usually want to grade his students' responses with some care, it is best to tape-record the students' answers so that they can be played back several times. Items of this sort are very time-consuming for all concerned, but they yield so much information for the attentive teacher that they are well worth the effort they require.

Chapter IV--GERMAN

PREPARING TEST ITEMS

4.1 INTRODUCTION

4.1.1 It is not easy to amass a large repertoire of effective test items. In addition to the difficult task of writing understandable instructions and credible distractors, the teacher must worry constantly about how the test fits what he has taught. In foreign-language work, it is often desirable for a test to sample all the language skills; on such a test testing devices must be used which are appropriate to each skill. We hope that the teacher will find the following sections helpful in making his own test items. This manual claims neither novelty nor completeness. In Chapter IV, we have tried, first, to be explicit about the essential elements of successful foreign-language tests and then to organize them into a useful working outline accompanied by concrete examples.

4.2 THE ESSENTIAL ELEMENTS OF A TEST

4.2.1 There are three basic sets of elements in any language test: (1) the language skills, (2) specific testing devices, and (3) the particular language areas to be tested.

4.2.2 The Language Skills

4.2.2.1 Teachers generally recognize four fundamental skills: (1) listening comprehension, (2) speaking, (3) reading, and (4) writing. In audio-lingual programs, listening comprehension is considered to have primacy over the others. The position taken is that one cannot speak, read or write meaningfully in a language without being able to understand that language. The other three skills are also interrelated, but it is not clear that speaking is as important for reading and writing as listening comprehension is for all of them. Even in the native language, the reading skill is often--perhaps always--more highly developed than the ability to write.

4.2.2.2 Tests must take into consideration which specific skills are to be evaluated. When the reading skill has been the main concern of the course, for instance, tests should not emphasize listening comprehension; in audio-lingual programs, which generally begin with listening comprehension and speaking, reading and writing should not be tested in the early stages of instruction.

4.2.2.3 Of the four skills, listening comprehension and reading are easiest to evaluate objectively. The difficulties inherent in testing writing and speaking are readily apparent. With listening and reading it is a comparatively simple matter to control the student's responses, but it is difficult to do so when the student is to react orally or in writing. Judgment of spoken and written responses involves much personal interpretation by the teacher; it is difficult to restrict the student to only one correct response. The burden of grading and administering oral-response tests makes them unwieldy; although written answers are somewhat easier to grade, they are just as difficult to control as spoken responses.

Having a number of teachers grade written responses may actually increase the degree of subjectivity, although some experts disagree.[1]

4.2.3 Testing Devices

4.2.3.1 The most commonly used devices in foreign-language testing are: (1) multiple-choice items, (2) completion items, (3) matching sets, (4) true/false items, (5) translation, (6) dictation, and items involving (7) expansion, (8) transformation, and (9) substitution. The last three are especially useful to test speaking and writing. Of the other devices, true/false items and translation are the ones most often debated. It is very difficult to phrase a true/false item so that it is clearly and unquestionably true or false but not transparently so. Translation has several shortcomings. Using translation from the native language into the foreign language to discover the student's control of specific problem areas (the subjunctive, for example, in German) is ineffective in cases where the student can find an alternate possibility (the infinitive, for example). Although translation from the foreign language to the native language is easier to control, its effectiveness is also somewhat limited because the student can usually find clues to the meaning of the phrase or word to be translated in language features (context, word order, etc.) other than the specific feature being tested. Therefore, unless the teacher can control quite strictly the context of a particular word or phrase, it is difficult to determine the student's actual command of the matter being tested. Extensive use of translation is obviously subject to attack in any curriculum that consciously emphasizes direct use and knowledge of the foreign language with as little recourse as possible to the student's native language. Perhaps translation is more accurately a specialized skill than a testing device. There is certainly no proof that the ability to understand, speak, read, and write a foreign language also guarantees the ability to translate freely between that language and one's own language. Indeed, in institutions with a sound foreign-language program designed for specialists, translation is often taught in a separate course to which students are admitted only after they have demonstrated a command of the other skills.

---

[1] See, for example, The Measurement of Writing Ability by Fred Godshalk, Frances Swineford and William Coffman, College Entrance Examination Board, New York (1966). The articulation of norms for written composition differs from language to language. For an illuminating discussion of procedures, see How the French Boy Learns to Write by Rollo Walter Brown, Harvard University Press (1915), reprinted by the National Council of Teachers of English (1963).

4.2.3.2 Multiple-choice items (i.e., choice of one correct answer from two or more possible answers) and matching sets are easy to control since variation in answers is strictly limited. To be effective, multiple-choice items must have distractors which are sufficiently similar to the correct choice to attract the student who does not really know what the correct answer is. Incorrect answers can reveal the individual student's problems. A wrong choice between Mutter and Mütter or wurden and würden in German, for example, may mean that the student needs more practice in distinguishing short *u* from short *ü*. By increasing the number of possible answers, the teacher can gain more information about the individual student's weaknesses and needs.

4.2.3.4 There are two ways of constructing *matching sets*: (1) the same number of items may be supplied in each of two columns, or (2) more items may appear in the column from which the 'answers' are to be selected, thereby leaving a residue when the answers have been chosen. In either case, matching is a type of interlocking multiple choice. Guessing, however, can be a factor in both matching and multiple-choice items. A formula for penalizing random choices can be used; it should be explained to the students in an effort to discourage them from making irresponsible guesses.

4.2.3.5 *Expansion items* test the student's knowledge of the construction of sentences in the foreign language. The student may be asked to expand a sentence in a number of different ways. In the following example, he is being tested on his knowledge of the proper word-order of declarative sentences that begin with an adverb:

Instructions: Insert *at the beginning* of each of the following sentences the word you are given:

Sentence No. 1: Ich lese das Buch.

Insert: Jetzt

Answer: Jetzt lese ich das Buch.

Sentence No. 2: Wir fahren nach München.

Insert: Am Mittwoch

Answer: Am Mittwoch fahren wir nach München.

4.2.3.6 Transformation items are particularly suited to evaluating writing. The student is given instructions and a model; for example, he is given a sentence with a verb in the present tense to be changed to a past tense. A series of sentences follows with the verb in the present tense which the student is to change to the past. Transformation items can be written to test more complex matters such as making active sentences passive, inserting subordinate clauses into simple sentences to make complex sentences, etc. It is always advisable to furnish a model item as part of the instructions.

4.2.3.7 In substitution items, the student is asked to replace a word, part of a word, or phrase with another:

Instructions:	Replace the underlined words with others which are equally correct.
Sentence:	Ich habe mir die Hände gewaschen.
Possible Answer:	Ich habe mir das Gesicht gewaschen.

4.2.3.8 Dictation is a well-known device in foreign-language instruction. However, it is difficult to grade straight dictation. The teacher must decide how to sort out and weight each student's individual errors. One way to avoid many problems is to give the student a partially filled-in answer sheet. As the teacher reads the dictation, the student writes in only what has been omitted from the text he has before him. The blanks may become more numerous as the student progresses in his ability to take dictation.

4.2.4 The Language Areas

4.2.4.1 The third basic set of elements involved in the preparation of foreign-language tests, the language areas to be tested, is the least clearly defined and by far the largest of the three. Opinions about the relative importance of various language areas differ widely.

4.2.5 Contrastive Analysis

4.2.5.1 The student's native language is carefully contrasted with the foreign language he is studying[2] because it is believed that the language learner will make numerous predictable errors when he incorrectly transfers the habits he has acquired in his native language to the language he is studying. In other words, the learner tends to express himself in the foreign language in terms of the rules of the language he already knows. He will automatically assume that whatever is not clearly different in the foreign language is the same as in his own language. A thorough knowledge of the differences that exist between the two languages can therefore help the teacher to anticipate problems.

4.2.5.2 A foreign-language test could well consist largely of items on the areas where the two languages differ. If the teacher knows, for example, that German grammatical gender is different from English gender, he can expect gender to be a problem for the English-speaking learner and may therefore devote a large proportion of his testing to gender in the early stages of instruction.

4.2.5.3 Wherever the native language of the student has a feature which is lacking or simpler in the foreign language, the problem is relatively easy to solve. Where German has nicht wahr?, for example, English has a far more complicated set of expressions. As a result, the English-speaking learner of German has to learn one invariable expression while the German-speaking student of English has to learn to use correctly the set of English expressions aren't we, couldn't we, etc.

---

[2] Cf. the Contrastive Structure Series (German-English) published by the University of Chicago Press, Charles A. Ferguson, General Editor.

4.2.5.4 Both the contrasts and the ways in which we state them depend on the theoretical premises we employ in our contrastive analysis. We suggest the following outline as a good guide to the arrangement of basic language matters:

I. Phonology (Sound System)

  a. phonemes (distinctive classes of sounds)

  b. distribution of phonemes (including diphthongs and consonant clusters)

  c. intonation (pitch changes in the voice to indicate statements, questions and emotional coloring)

II. Morphology (Formation of Words)

  a. morphemes (the basic units of grammar)

  b. inflection (changes in words according to number, person, case, etc.)

  c. derivation of words (formation of new words by compounding or by the addition of prefixes and suffixes)

III. Syntax (Arrangements of Words in Larger Constructions)

  a. phrases

  b. clauses

  c. sentences

4.2.6 Vocabulary

The size and nature of a student's vocabulary is obviously important in language teaching. Various word counts exist,[3] but there is surprisingly little agreement about what words should be taught at each level of language instruction. However, since the teacher must test vocabulary, he should select vocabulary items as shrewdly as he can to test those items that seem most useful. For purposes of classification, idioms should be included in 'vocabulary'. It is also often useful to group words into families or general categories.

4.2.7 Style

Speakers of German make systematic use of a number of different styles of discourse of varying degrees of formality and impersonality, both orally and in writing. Some distinctions in style affect even basic grammatical forms. Such is the case with the forms of polite and familiar address, for example. At least the most important distinctions of style are treated in most language texts and should be tested.

4.2.8 Literature

Literature is generally considered an essential part of foreign-language instruction. In audio-lingual programs literature is not ignored; rather, it is not treated until students can read with some ease and with sensitivity to literary values. The choice of specific literary selections varies, of course, from teacher to teacher and from class to class.

---

[3] An interesting study has been done for English at Brown University (see the forthcoming publication, Computational Analyses of Present Day American English, edited by Henry Kucera and W. N. Francis). In a text of 1,014,235 words, there were only 50,457 different words; 42,540 of these occurred only ten times or less.

4.2.9 Culture

There is no generally accepted, useful definition of culture in the context of foreign language teaching. Culture is understood either in a narrow sense, meaning understanding of the literary and artistic values general among the educated speakers of a language, or it is extended to include the whole way of life and the history of a people, the geography of the land they live in and their political, religious and social institutions. Teachers seem to agree, however, that culture is an essential part of language instruction and that they should test their students on it.

4.2.10 A Brief List of Language Areas

The following list of language areas is offered:

(1) distinctive sound classes (phonology) (e.g., the vowel sounds in Haar, Heer, hier, etc.);

(2) diphthongs and other combinations of vowels (e.g., laufe, läuft, ruhe);

(3) consonant clusters (e.g., /mpf/ in Strumpf);

(4) statement intonation;

(5) question intonation;

(6) basic grammar units (e.g., von/aus, sein/stehen/liegen);

(7) inflectional procedures (e.g., declension of adjectives);

(8) derivation (e.g., formation of compounds, etc.);

(9) phrasal syntax (e.g., the positioning of prepositions: nach Berlin, meiner Meinung nach);

(10) clausal syntax (e.g., word order in subordinate and coordinate clauses);

(11) sentence syntax (e.g., sequence of tenses);

(12) vocabulary (including both single words and idioms like Hunger, Hunger haben);

(13) style (e.g., very formal expository prose);

(14) literature (names of well-known writers, reading selections);

(15) culture (accepted ways of speaking in specified social situations, ways of addressing officials, etc.).

The list is open; additional features can be added to it and the ones given can be made more detailed.

4.3 WORKING OUTLINES FOR TEST WRITING

The skills and testing devices have been arranged on the following chart:

Chart No. 1: Skills and Testing Devices

Skills	Testing Devices										
	MC	CO	MA	T/F	TR	DIC	EX	T	SUB	RA	EC
Listening Comprehension	X		X		X	X		X	X		
Speaking	X	X			X		X	X	X	X	
Reading	X		X	X	X		X	X		X	X
Writing	X	X			X	X	X	X	X		X

Key:

MC - multiple choice
CO - completion
MA - matching
T/F - true/false
TR - translation
DIC - dictation
EX - expansion
T - transformation
SUB - substitution
RA - reading aloud
EC - extended composition
X - effective use

Here are some useful general observations on the appropriateness of particular testing devices to specific skills:

(1) Generally speaking, completion items are hard to control. In order to assure objectivity, completion items must be restricted to absolutely unambiguous possibilities. (Visuals can be useful for this purpose.)

(2) Ambiguity is a major problem in true/false questions. The teacher should be especially cautious in his use of ungrammatical sentences which the student is to identify as "incorrect." Just as correct stimuli appear to reinforce learning, incorrect stimuli may create uncertainty where the student is well prepared.

(3) The use of translation to test all four skills is open to question. Problems arise in restricting translation so as to test only specific points. Translation can also be considered a distinct skill; as such, it, too, can be tested by a number of devices, such as multiple choice, completion, matching and true/false items.

(4) Expansion and transformation items are useful devices for testing speaking and writing, but they are inappropriate devices for testing the receptive skills of listening and reading.

It is reasonable to suppose that some testing devices are more suitable than others to pupils of different ages. In the absence of evidence from controlled experimentation, the individual teacher must rely on his own judgment.

4.3.1 Chart No. 2, combining language areas and skills, is a check list for the preparation of test items. By filling it with actual test items of different kinds, the teacher can achieve a broad coverage of language problems. The following samples illustrate item composition. Each item is numbered in cross reference to Chart No. 2.

Chart No. 2: Language Areas and Skills

Language Areas	Skills			
	Listening Comprehension	Speaking	Reading	Writing
Sound Classes (Phonemes)	(1)	(4)	(4)	
Orthography	(14)	(3)	(3)	(14)
Consonant Clusters		(4)	(4)	
Grammar Units (Morphemes)	(2)	(2)	(6)	
Inflection	(5)	(5)		
Syntax: phrases			(11)	(11)
Syntax: clauses			(10)	(10)
Syntax: sentences			(9)(10) (12)(13)	(9)(10) (12)(13)
Vocabulary			(7)(8)	
Style			(13)	(13)
Literature			(15)	
Culture			(16)	

4.3.2 Sample Test Items

(1) Language Area: Sound classes (Phonemes)

Skill: Listening comprehension

Testing Device: Multiple choice

Instructions: You will hear an utterance. On your answer sheet circle the number of the utterance you heard.

Student hears: eine Kirche

Student sees:

1. eine Küche
2. eine Kirche
3. eine Kirsche

Answer: 2

Comment:

This item tests sound discrimination, in particular, the distinction between i and ü, and ch and sch; it also involves the reading skill.

(2) Language Area: Basic grammar

Skills: Listening comprehension/Speaking

Testing Device: Substitution/Transformation

Instructions: Repeat the sentences making substitutions required as illustrated in the example.

Example:

Student hears: Der Vater spricht mit seinem Sohn. (Mutter)

Answer: Die Mutter spricht mit ihrem Sohn.

Student hears:

1. Das Kind ruft seine Mutter. (seine Schwester)
2. Der Student macht seine Arbeit. (die Studenten)
3. Sie schreibt mit ihrem Kugelschreiber. (wir)

Answers:

1. Seine Schwester ruft ihre Mutter.
2. Die Studenten machen ihre Arbeit.

3. Wir schreiben mit unserem Kugelschreiber.
or: Wir schreiben mit unseren Kugelschreibern.

Comment:

This item tests noun-pronoun congruency (case, number, gender).

(3) Language Area: Orthography

Skills: Reading/Speaking

Testing Device: Reading aloud

Instructions: Read the following sentences aloud:

1. Das ist ein Schiff.
2. Er hat eine tiefe Stimme.
3. Er verlor die Sprache.
4. Er hat ein Skelett.

Answers: (Scorer scores only pronunciation of the underlined portions.)

1. Schiff
2. Stimme
3. Sprache
4. Skelett

Comment:

This item tests association between sound and written symbol.

(4) Language Areas: Sound classes/Consonant clusters

Skills: Reading/Speaking

Testing Device: Reading aloud

Instructions: Read the following sentences aloud:

1. Er hat sein Buch vergessen.
2. Das Obst war in der Tüte.
3. Wir fahren mit der Strassenbahn.
4. Er kauft ein Pfund Äpfel.

Answers: (Scorer scores only the pronunciation of the underlined portions.)

1. Buch
2. Obst, Tüte
3. Strassenbahn
4. Pfund, Äpfel

Comment:

This item tests association between sound and written symbol and the pronunciation of difficult consonant clusters and vowels.

(5) Language Area: Inflection

Skills: Listening comprehension/Speaking

Testing Device: Transformation

Instructions: Make the following sentences plural.

Example:

The student hears: Das Auto fährt langsam.

The student should answer: Die Autos fahren langsam.

The student hears:

1. Der Vogel singt.
2. Das Kind ist eingeschlafen.
3. Noch immer hört man das komische Geräusch.

The student should answer:

1. Die Vögel singen.
2. Die Kinder sind eingeschlafen.
3. Noch immer hört man die komischen Geräusche.

Comment:

This item tests the transformation from singular to plural forms (article, adjective, noun, finite verb).

(6) Language Area: Basic grammar

Skill: Reading

Testing Device: Multiple choice

Instructions: Circle the pronoun which correctly replaces the underlined word.

Ursula gab Helmut das Buch.

a. ihr
b. ihn
c. ihm
d. ihnen

Answer: c

(7) Language Area: Vocabulary

Skill: Reading

Testing Device: Multiple choice

Instructions: Choose the item which can be substituted for the underlined word without changing the meaning of the sentence.

Das bekommt man hier nicht.

a. kommt
b. kauft
c. kriegt
d. kennt

Answer: c

Comment:

This item tests synonyms.

(8) Language Area: Vocabulary

Skill: Reading

Testing Device: Multiple choice

Instructions: Choose the best item to complete the following sentences:

1. Die Strassen sind so schlecht, dass die Luft ganz ____________ ist.

   a. regnerisch
   b. neblig
   c. staubig
   d. verschneit

2. Reichtum ist das Gegenteil von __________

   a. Demut
   b. Hochmut
   c. Armut
   d. Mut

Answers:

1. c
2. c

Comment:

These items test (1) adjectives and (2) antonyms.

(9) Language Area: Sentence syntax

Skills: Writing/Reading

Testing Device: Transformation

Instructions: Write a complete sentence using the following words in the order given. Make all necessary changes and add what is needed.

Example: Er / reisen / oft / in / nächst / Stadt

Answer: Er reist (reiste) oft in die nächste Stadt.

Mein / jung / Sohn / fahren / gern / sein / Fahrrad

Answer:

Mein junger (jüngster, jüngerer) Sohn fährt gern mit seinem Fahrrad.

Comment:

This item tests grammar, in particular, a sentence synthesis from pronominal, adverbial, adjectival, verbal, and nominal base forms; it also involves transformation, and the reading skill.

(10) Language Areas: Sentence syntax/Clausal syntax

Skills: Writing/Reading

Testing Device: Transformation

Instructions: Rewrite the following two sentences as a single sentence starting with the underlined words. Add any words that may be necessary.

Der Mann rief seinen Hund. Er wollte nach Hause gehen.

Answer:

Der Mann rief seinen Hund, weil er nach Hause gehen wollte.

or: Der Mann, der nach Hause gehen wollte, rief seinen Hund.

or: Der Mann rief seinen Hund, denn er wollte nach Hause gehen.

(11) Language Area: Phrasal syntax

Skills: Writing/Reading

Testing Device: Expansion

Instructions: Rewrite the sentence, incorporating into it the adjective you are given:

Example: Ich kaufe die Karte. (billig)

Answer: Ich kaufe die billige Karte.

Ich ziehe mir den Rock an. (dunkel)

Answer:

Ich ziehe mir den dunklen Rock an.

(12) Language Area: Sentence syntax

Skills: Writing/Reading

Testing Device: Transformation

Instructions: For each of the following, write a question asking for the information given in the underlined part of the sentence.

Example: Die Bücher liegen unter dem Tisch.

Answer: Worunter liegen die Bücher?
or: Wo liegen die Bücher?

1. Sie denkt an die schönen Feiertage.
2. Du sollst mit Tinte schreiben.
3. Sie gehen heute mit Otto spazieren.

Answers:

1. An was denkt sie?
or: Woran denkt sie?

2. Mit was sollst du schreiben?
or: Womit sollst du schreiben?

3. Mit wem gehen sie heute spazieren?

Comment:

There may be more than one possible question format in each instance.

(13) Language Areas: Sentence syntax/Style

Skills: Writing/Reading

Testing Device: Extended composition

Instructions: Write a short paragraph in German using the answers to each of the following questions as a guide.

Das Telefongespräch

1. Wer rief Sie an?
2. Warum wollte der Betreffende mit Ihnen sprechen?
3. Welche Auskunft gaben Sie ihm?

Comment:

This item tests grammar and vocabulary, and sentence formation within the complex context of an extended prose composition.

(14) Language Area: Orthography

Skills: Listening comprehension/Writing

Testing Device: Dictation

Instructions: Write the following words:

Student hears:

Meister
Münster
Muster
Mast
Messe

Strasse
Gasse
Nase
Glas
nahe
Sahne

Comment:

This item tests spelling; in particular, it tests the correct use of consonants and vowel symbols in combination to indicate vowel quantities.

(15) Language Area: Literature

Skill: Reading

Testing Device: Multiple choice

Instructions: Who wrote the following? Circle the letters corresponding to the correct answers:

1. Geist der Goethezeit

   a. Fairley
   b. Korff
   c. Morris
   d. Bielschafsky

2. Lenore

   a. Schiller
   b. Klinger
   c. Bürger
   d. Goethe

Answers:

1. b
2. c

Comment:

This item tests retention of specific facts of literary history.

(16) Language Area: Culture

Skill: Reading

Testing Device: Multiple choice

Instructions: Circle the letter corresponding to the correct choice.

1. Approximately what dates delimit the period of Old High German?

   a. 300 - 750
   b. 750 - 1100
   c. 1100 - 1350

2. The German Federal Republic is made up of several:

   a. Staaten
   b. Länder
   c. Provinzen

Answers:

1. b
2. b

Comment:

These items test specific items of culture.

Chapter IV--ITALIAN

PREPARING TEST ITEMS

4.1 INTRODUCTION

4.1.1 It is not easy to amass a large repertoire of effective test items. In addition to the difficult task of writing understandable instructions and credible distractors, the teacher must worry constantly about how the test fits what he has taught. In foreign-language work, it is often desirable for a test to sample all the language skills; on such a test, testing devices must be used which are appropriate to each skill. We hope that the teacher will find the following sections helpful in making his own test items. This manual claims neither novelty nor completeness. In Chapter IV, we have tried, first, to be explicit about the essential elements of successful foreign-language tests and then to organize them into a useful working outline accompanied by concrete examples.

4.2 THE ESSENTIAL ELEMENTS OF A TEST

4.2.1 There are three basic sets of elements in any language test: (1) the language skills, (2) specific testing devices, and (3) the particular language areas to be tested.

4.2.2 The Language Skills

4.2.2.1 Teachers generally recognize four fundamental skills: (1) listening comprehension, (2) speaking, (3) reading, and (4) writing. In audio-lingual programs, listening comprehension is considered to have primacy over the others. The position taken is that one cannot speak, read or write meaningfully in a language without being able to understand that language. The other three skills are also interrelated, but it is not clear that speaking is as important for reading and writing as listening comprehension is for all of them. Even in the native language, the reading skill is often--perhaps always--more highly developed than the ability to write.

4.2.2.2 Tests must take into consideration which specific skills are to be evaluated. When the reading skill has been the main concern of the course, for instance, tests should not emphasize listening comprehension; in audio-lingual programs, which generally begin with listening comprehension and speaking, reading and writing should not be tested in the early stages of instruction.

4.2.2.3 Of the four skills, listening comprehension and reading are easiest to evaluate objectively. The difficulties inherent in testing writing and speaking are readily apparent. With listening and reading it is a comparatively simple matter to control the student's responses, but it is difficult to do so when the student is to react orally or in writing. Judgment of spoken and written responses involves much personal interpretation by the teacher; it is difficult to restrict the student to only one correct response. The burden of grading and administering oral-response tests makes them unwieldy; although written answers are somewhat easier to grade, they are just as difficult to control as spoken responses. Having a number of teachers grade written responses may actually increase the degree of subjectivity, although some experts disagree.[1]

4.2.3 Testing Devices

4.2.3.1 The most commonly used devices in foreign-language testing are: (1) multiple-choice items, (2) completion items, (3) matching sets, (4) true/false items, (5) translation, (6) dictation, and items involving (7) expansion, (8) transformation, and (9) substitution. The last three are especially useful to test speaking and writing. Of the other devices, true/false items and translation are the ones most often debated. It is very difficult to phrase a true/false item so that it is clearly and unquestionably true or false but not transparently so. Translation has several shortcomings. Using translation from the native language into the foreign language to discover the student's control of specific problem areas (the subjunctive, for example, in Italian) is ineffective in cases where the student can find an alternate possibility (the infinitive, for example). Although translation from the foreign language to the native language is easier to control, its effectiveness is also somewhat limited because the student can usually find clues to the meaning of the phrase or word to be translated in language features (context, word order, etc.) other than the specific feature being tested. Therefore, unless the teacher can control quite strictly the context of a particular word or phrase, it is difficult to determine the student's actual command of the matter being tested. Extensive use of translation is obviously subject to attack in any curriculum that consciously emphasizes direct use and knowledge of the foreign language with as little recourse as possible to the student's native language. Perhaps translation is more accurately a specialized skill than a testing device. There is certainly no proof that the ability to understand, speak, read, and write a foreign language also guarantees the ability to translate freely between that language and one's own language. Indeed, in institutions with a sound foreign-language program designed for specialists, translation is often taught in a separate course to which students are admitted only after they have demonstrated a command of the other skills.

---

[1] See, for example, The Measurement of Writing Ability by Fred Godshalk, Frances Swineford and William Coffman, College Entrance Examination Board, New York (1966). The articulation of norms for written composition differs from language to language. For an illuminating discussion of procedures, see How the French Boy Learns to Write by Rollo Walter Brown, Harvard University Press (1915), reprinted by the National Council of Teachers of English (1963).

4.2.3.2 Multiple-choice items (i.e., choice of one correct answer from two or more possible answers) and matching sets are easy to control since variation in answers is strictly limited. To be effective, multiple-choice items must have distractors which are sufficiently similar to the correct choice to attract the student who does not really know what the correct answer is. Incorrect answers can reveal the individual student's problems. A wrong choice between fato and fatto or copia and coppia in Italian, for example, may mean that the student needs more practice in distinguishing single from double consonants. By increasing the number of possible answers, the teacher can gain more information about the individual student's weaknesses and needs.

4.2.3.4 There are two ways of constructing matching sets: (1) the same number of items may be supplied in each of two columns, or (2) more items may appear in the column from which the 'answers' are to be selected, thereby leaving a residue when the answers have been chosen. In either case, matching is a type of interlocking multiple choice. Guessing, however, can be a factor in both matching and multiple-choice items. A formula for penalizing random choices can be used; it should be explained to the students in an effort to discourage them from making irresponsible guesses.

4.2.3.5 Expansion items test the student's knowledge of the construction of sentences in the foreign language. The student may be asked to expand a sentence in a number of different ways. In the following example, he is being tested on his knowledge of the positioning of adjectival modifiers in noun phrases:

Instructions:	In each of the following sentences insert the word you are given:
Sentence No. 1:	Le ragazze sono venute.
Insert:	belle
Answer:	Le belle ragazze sono venute.

Sentence No. 2: Gli studenti hanno studiato.

Insert: intelligenti

Answer: Gli studenti intelligenti hanno studiato.

4.2.3.6 *Transformation items* are particularly suited to evaluating writing. The student is given instructions and a model; for example, he is given a sentence with a verb in the present tense to be changed to a past tense. A series of sentences follows with the verb in the present tense which the student is to change to the past. Transformation items can be written to test more complex matters such as making active sentences passive, inserting subordinate clauses into simple sentences to make complex sentences, etc. It is always advisable to furnish a model item as part of the instructions.

4.2.3.7 In *substitution items*, the student is asked to replace a word, part of a word, or phrase with another:

Instructions: Replace the underlined words with others which are equally correct.

Sentence: Mi sono lavato *le* *mani*.

Possible Answer: Mi sono lavato la faccia.

4.2.3.8 *Dictation* is a well-known device in foreign-language instruction. However, it is difficult to grade straight dictation. The teacher must decide how to sort out and weight each student's individual errors. One way to avoid many problems is to give the student a partially filled-in answer sheet. As the teacher reads the dictation, the student writes in only what has been omitted from the text he has before him. The blanks may become more numerous as the student progresses in his ability to take dictation. Dictation has, however, only limited utility in teaching Italian.

4.2.4 The Language Areas

4.2.4.1 The third basic set of elements involved in the preparation of foreign-language tests, the language areas to be tested, is the least clearly defined and by far the largest of the three. Opinions about the relative importance of various language areas differ widely.

4.2.5 Contrastive Analysis

4.2.5.1 The student's native language is carefully contrasted with the foreign language he is studying[2] because it is believed that the language learner will make numerous predictable errors when he incorrectly transfers the habits he has acquired in his native language to the language he is studying. In other words, the learner tends to express himself in the foreign language in terms of the rules of the language he already knows. He will automatically assume that whatever is not clearly different in the foreign language is the same as in his own language. A thorough knowledge of the differences that exist between the two languages can therefore help the teacher to anticipate problems.

4.2.5.2 A foreign-language test could well consist largely of items on the areas where the two languages differ. If the teacher knows, for example, that Italian grammatical gender is different from

---

[2] Cf. the Contrastive Structure Series (Italian-English) published by the University of Chicago Press, Charles A. Ferguson, General Editor.

English gender, he can expect gender to be a problem for the English-speaking learner and may therefore devote a large proportion of his testing to gender in the early stages of instruction.

4.2.5.3 Wherever the native language of the student has a feature which is lacking or simpler in the foreign language, the problem is relatively easy to solve. Where Italian has a non è vero, for example, English has a far more complicated set of expressions. As a result, the English-speaking learner of Italian has to learn one invariable expression while the Italian student of English has to learn to use correctly the set of English expressions aren't we, couldn't we, etc.

4.2.5.4 Both the contrasts and the ways in which we state them depend on the theoretical premises we employ in our contrastive analysis. We suggest the following outline as a good guide to the arrangement of basic language matters:

I. Phonology (Sound System)

   a. phonemes (distinctive classes of sounds)

   b. distribution of phonemes (including diphthongs and consonant clusters)

   c. intonation (pitch changes in the voice to indicate statements, questions and emotional coloring)

II. Morphology (Formation of Words)

   a. morphemes (the basic units of grammar)

   b. inflection (changes in words according to number, person, case, etc.)

   c. derivation of words (formation of new words by compounding or by the addition of prefixes and suffixes)

III. Syntax (Arrangements of Words in Larger Constructions)

  a. phrases

  b. clauses

  c. sentences

4.2.6 Vocabulary

The size and nature of a student's vocabulary is obviously important in language teaching. Various word counts exist,[3] but there is surprisingly little agreement about what words should be taught at each level of language instruction. However, since the teacher must test vocabulary, he should select vocabulary items as shrewdly as he can to test those items that seem most useful. For purposes of classification, idioms should be included in 'vocabulary'. It is also often useful to group words into families or general categories.

4.2.7 Style

Italians make systematic use of a number of different styles of discourse of varying degrees of formality and impersonality, both orally and in writing. Some distinctions in style affect even basic grammatical forms. Such is the case with the forms of polite and familiar address in Italian, a language which has more than one equivalent to the English "you." At least the most important distinctions of style are treated in most language texts and should be tested. For example, the student should be taught to recognize the stylistic incongruity of a combination like ciao, professore as well as he does in its English equivalent so long, professor!

---

[3] An interesting study has been done for English at Brown University (see the forthcoming publication, Computational Analyses of Present Day American English, edited by Henry Kucera and W. N. Francis). In a text of 1,014,235 words, there were only 50,457 different words; 42,540 of these occurred only ten times or less.

4.2.8 Literature

Literature is generally considered an essential part of foreign-language instruction. In audio-lingual programs literature is not ignored; rather, it is not treated until students can read with some ease and with sensitivity to literary values. The choice of specific literary selections varies, of course, from teacher to teacher and from class to class.

4.2.9 Culture

There is no generally accepted, useful definition of culture in the context of foreign language teaching. Culture is understood either in a narrow sense, meaning understanding of the literary and artistic values general among the educated speakers of a language, or it is extended to include the whole way of life and the history of a people, the geography of the land they live in and their political, religious and social institutions. Teachers seem to agree, however, that culture is an essential part of language instruction and that they should test their students on it.

4.2.10 A Brief List of Language Areas

The following list of language areas is offered with examples:

(1) distinctive sound classes (phonology) (e.g., the vowel sounds da, de', di, do);

(2) diphthongs and longer combinations of vowels (e.g., dai, dei, vuoi);

(3) consonant clusters (e.g., /sb/ in sbaglio);

(4) statement intonation (e.g., the falling pitch levels in Italian):

(5) question intonation (e.g., Vai a Roma? versus Vai a Roma.);

(6) basic grammar units (e.g., di/da, essere/stare);

(7) inflectional procedures (e.g., verbal inflection);

(8) derivation (e.g., formation of diminutives, etc.);

(9) phrasal syntax (e.g., the positioning of adjectives: un pover' uomo, un uomo povero);

(10) clausal syntax (e.g., use of clauses introduced by che, etc.);

(11) sentence syntax (e.g., sequence of tenses);

(12) vocabulary (including both single words and idioms like fame, aver fame);

(13) style (e.g., use of egli in written Italian);

(14) literature (names of well-known writers, reading selections);

(15) culture (accepted ways of speaking in specified social situations, ways of addressing officials, etc.).

The list is open; additional features can be added to it and the ones given can be made more detailed.

4.3 WORKING OUTLINES FOR TEST WRITING

The skills and testing devices have been arranged on the following chart:

Chart No. 1: Skills and Testing Devices

Skills	Testing Devices								
	MC	CO	MA	T/F	TR	DIC	EX	T	SUB
Listening Comprehension	X		X		X				
Speaking	X	X			X		X	X	X
Reading	X		X	X	X				
Writing	X	X			X	X	X	X	X

Key:
MC - multiple choice
CO - completion
MA - matching
T/F - true/false
TR - translation
DIC - dictation
EX - expansion
T - transformation
SUB - substitution
X - effective use

An examination of the chart leads to some general observations on the appropriateness of particular testing devices to specific skills:

(1) Generally speaking, completion items are hard to control. In order to assure objectivity, completion items must be restricted to absolutely unambiguous possibilities. (Visuals can be useful for this purpose).

(2) Ambiguity is a major problem in true/false questions. The teacher should be especially cautious in his use of ungrammatical sentences which the student is to identify as "incorrect." Just as correct stimuli appear to reinforce learning, incorrect stimuli may

create uncertainty where the student is well prepared.

(3) The use of translation to test all four skills is open to question. Problems arise in restricting translation so as to test only specific points. Translation can also be considered a distinct skill; as such, it, too, can be dested by a number of devices, such as multiple choice, completion, matching and true/false items.

(4) Expansion and transformation items are useful devices for testing speaking and writing, but they are inappropriate devices for testing the receptive skills of listening and reading.

It is reasonable to suppose that some testing devices are more suitable than others to pupils of different ages. In the absence of evidence from controlled experimentation, the individual teacher must rely on his own judgment.

4.3.1 Chart No. 2, combining language areas and skills, is a check list for the preparation of test items. By filling it with actual test items of different kinds, the teacher can achieve a broad coverage of language problems. The following samples illustrate item composition. Each item is numbered in cross reference to Chart No. 2.

Chart No. 2: Language Areas and Skills

Language Areas	Skills			
	Listening Comprehension	Speaking	Reading	Writing
Sound Classes (Phonemes)	(1)			
Diphthongs (and longer chains of vowels)		(7)		
Consonant Clusters	(2)			
Statement Intonation		(8)		
Question Intonation	(3)			
Grammar Units (Morphemes)				(15)
Inflection		(9)(4)		(4)
Derivation				(16)
Syntax: phrases			(11)	
Syntax: clauses	(5)	(5)		(5)
Syntax: sentences			(12)	
Vocabulary	(6)		(13)	
Style				(17)
Literature			(18)	
Culture		(10)	(14)	

4.3.2 Sample Test Items

(1) *Language Area:* Sound classes (Phonemes)

Skill: Listening comprehension

Testing Device: Multiple choice

Instructions:

Select the word you will hear from the following possibilities: A. callo, B. calo, C. caglio

(Teacher then says: calo.)

Comment:

Contrastive analysis of Italian and English reveals that the Italian phonemic contrast between single and double consonants can be difficult to learn for speakers of English. Two of the three possible answers (*callo* and *calo*) exemplify this contrast. The third possibility (*caglio*) serves as a plausible distractor. Other items of this kind can be worked out easily: /p/ - /pp/ (cópia - coppia), /s/ - /ss/ (casa - cassa), etc.

(2) *Language Area:* Consonant clusters

Skill: Listening comprehension

Testing Device: Matching

Instructions:

Match the words in column A with those having the same initial sound in column B. Model words: *specchio*, *sbaglio*. (The teacher reads these two words aloud carefully.)

Column A	Column B
1. smetto	a. sblocco
2. stesso	b. spesso
3. snello	

Answers: 1-a, 2-b, 3-a.

Comment:

It is well known that English-speaking students have difficulty combining [z] with voiced consonants in word-initial position in Italian. The matching device enables the teacher to determine if the student has been able to perceive the correct pattern of distribution of [s] and [z].

(3) Language Area: Question intonation

Skill: Listening comprehension

Testing Device: Multiple choice

Instructions:

Select the sentence that is most likely to follow the model sentence in normal conversation. Model sentence: Enrico dorme? (pronounced with the normal rising question intonation in Italian). Choices for answers are: A. No, si è alzato. B. Non mi alzo. C. Lo saprò. D. Hai fame? Answer: A.

Comment:

A common question intonation pattern in Italian does not require word inversion. In this item the student is tested primarily on whether he is able to distinguish a question from a statement.

(4) Language Area: Inflection of verbs

Skills: Writing/Speaking

Testing Device: Completion

Instructions:

Complete the sentences with the appropriate future form of the indicated verbs.

1. Giovanni _________ la lezione. (studiare)
2. I miei amici _________ a scuola. (andare)
3. Noi lo _________ stasera. (vedere)

Answers: 1. studiera
2. andranno
3. vedremo

Comment:

Items of this kind test the student's ability to produce correct forms in specific contexts. It is reasonable to suppose that the student could speak whatever he can write, although hesitation might be more obvious in speech.

(5) Language Area: Clausal syntax

Skills: Listening comprehension/Writing/Speaking

Testing Device: Multiple choice

Instructions:

Choose the sequence that correctly completes the sentence that you will hear. Teacher says: Era impossibile che . . . Students then choose from the following: A. venga, B. venisse, C. venite, D. è venuto.

Comment:

Although all the choices are real verb forms, only B is grammatically correct in this context. The student must both hear the stem accurately and choose the form to complete it.

(6) Language Area: Vocabulary

Skill: Listening comprehension

Testing Device: True/False

Instructions:

Listen to the following statement and indicate whether it is true or false: Andare è sinonimo di venire. A. True, B. False.

Comment:

Sets of antonyms or synonyms should be clear, especially in elementary and intermediate instruction. Questions of this sort can be gradually made more subtle as the students' knowledge of Italian grows. The difficulty of using true/false is evident. Nevertheless, it is a useful device to test the student's range of vocabulary and his knowledge of families of words.

(7) Language Areas: Diphthongs and Triphthongs

Skill: Speaking

Testing Device: Transformation

Instructions:

Following the pattern of sapere/sai, pronounce the tu form of the following verbs: essere, andare, avere, potere, volere.

Comment:

Since the teacher is concerned here with how well the student can pronounce the diphthongs /ei/ and /ai/ and the triphthong /uoi/, the grammatical change is made as obvious as possible. The student may say the forms directly to the teacher or record them on tape.

(8) Language Area: Statement intonation

Skill: Speaking

Testing Device: Transformation

Instructions:

On the model of: Gli piace il vino?/Gli piace il vino., change the following questions to statements: Studia l'italiano? Vai a Roma?

Comment:

Grammatical change is avoided in order to focus on the student's ability to produce the statement intonation.

(9) Language Area: Inflection

Skill: Speaking

Testing Device: Transformation

Instructions:

Change the verb in the following sentences to the future tense. Model: Studio l'italiano. - Studierò l'italiano.

1. Canta la canzone.
2. Fa il suo lavoro.
3. Leggiamo il nuovo libro.

Answers:
1. Canterà la canzone.
2. Farà il suo lavoro.
3. Leggeremo il nuovo libro.

Comment:

The sentences may be given to the student orally, or he may read them from a sheet and say the answers aloud. If the teacher varies the person and the verb (as in this case), two possible errors must be anticipated: tense and person. Items of this kind can be more complicated; for instance, they can involve correct sequence of tenses in subordinate clauses, etc.

(10) Language Area: Culture (social institutions)

Skill: Writing

Testing Device: Completion

Instructions:

Write a short and socially appropriate rejoinder to the following:

1. Diamoci del tu.
2. Sono lieta di fare la Sua conoscenza.
3. Buon pranzo!

Possible answers:
1. D'accordo.
2. Il piacere è mio, Signora.
3. Grazie, altrettanto.

Comment:

This item involves only one facet of the many-sided composite that is culture. The student is being tested on his ability to react to expressions that fit commonly encountered situations. Some latitude in the answer will have to be allowed by the teacher (omission of "Signora" in answer to No. 2, for example), but no answer should be allowed which would be inappropriate in the social circumstances implied by the context.

(11) Language Area: Phrasal syntax

Skill: Reading

Testing Device: Matching

Instructions:

Match each relative clause in Column B with its antecedent in Column A:

Column A	Column B
1. Dov'è il signore	a. la quale voleva comprare il caffè?
2. E'venuta la signora	b. il quale voleva parlarmi?

Answers: 1-b, 2-a.

Comment:

The student is tested on his ability to maintain gender across phrasal boundaries.

(12) Language Area: Sentence syntax

Skill: Reading

Testing Device: Matching

Instructions:

Match the clauses in Column B with those of Column A to make coherent sentences:

Column A	Column B
1. Se avrò abbastanza danaro	a. ti avrei comprato un bel regalo.
2. Se avessi abbastanza danaro	b. ti comprerò un bel regalo.

Answers: 1-b, 2-a.

Comment:

The student is being tested on his ability to recognize appropriate sequences of tenses in complex sentences.

(13) Language Area: Vocabulary

Skill: Reading

Testing Device: Multiple choice

Instructions:

Select the expression which best completes the question.

1. Sai ____________________?

    a. il professor Marino
    b. la Francia
    c. nuotare

2. Conosci ____________________ bene?

    a. l'Italia
    b. pattinare
    c. cucire

Answers: 1-c, 2-a.

Comment:

The distinction between sapere and conoscere is being tested in context.

(14) Language Area: Culture (Music)

Skill: Reading

Testing Device: Matching

Instructions:

Match the operas in Column A with the names of their composers in Column B.

Column A	Column B
1. Lucia di Lammermoor	a. Pergolesi
2. La Serva Padrona	b. Puccini
3. Suor Angelica	c. Donizetti
4. l'Amico Fritz	d. Bellini
5. Norma	e. Mascagni

Answers: 1-c, 2-a, 3-b, 4-e, 5-d.

Comment:

Matching is a well established device to test factual knowledge. Other devices could have been used here; different types of items could render the recall of the facts more or less difficult. By using true/false items, we could present the student with the following:

1. Mascagni scrisse Lucia di Lammermoor. a. True, b. False.
2. Un'opera di Giacomo Puccini è Suor Angelica. a. True, b. False.

We could use completion items such as:

1. ______________ scrisse l'Amico Fritz.

(15) Language Area: Grammar units

Skill: Writing

Testing Device: Completion

Instructions:

Fill in the blanks with appropriate verb forms to indicate a simple past tense:

1. Le ragazze _________ venute.
2. I soldati si _________ seduti.
3. Giovanni _________ corso un miglio.

Answers: 1. sono, 2. sono, 3. ha

Comment:

The proper use of the auxiliary verbs is a major problem for students of Italian whose native language is English. The most difficult of these three items is the last one (Giovanni _________ corso un miglio.) because the verb correre takes the auxiliary verb essere when it is intransitive and avere when followed by a direct object.

(16) Language Area: Derivation

Skill: Writing

Testing Device: Completion

Instructions:

Fill in the correct derivative suffix in the following blanks:

1. Il mio fratell_______ non vuole andare a scuola.
   (My little brother doesn't want to go to school.)
2. Un buon giocat_______ guadagna spesso.
   (A good player often wins.)
3. Il pane_______ fa il pane.
   (The bread-maker makes bread.)

Answers: 1. -ino, 2. -ore, 3. -ttiere.

Comment:

The sentences must be chosen with care in order to insure that the student will be either completely right or completely wrong in his choice of derivative.

(17) Language Area: Style

Skill: Writing

Testing Device: Transformation

Instructions:

Change the following expressions to their informal equivalents:

1. Come si chiama Lei?
2. ArrivederLa.

3. Come stanno Loro?

Answers: 1. Come ti chiami? 2. Ciao. 3. Come state?

Comment:

These items test the student's ability to shift from formal to informal style. The choice of items must be carefully considered to take into account alternate informal expressions, such as the possibility of *arrivederci* for answer No. 2.

(18) Language Area: Culture

Skill: Reading

Testing Device: Matching

Instructions:

Match the following authors and works. Write the appropriate letters from the second column in the spaces provided before the authors' names.

_____ 1. Dante a. *Il Principe*

_____ 2. Machiavelli b. *I Promessi Sposi*

_____ 3. Manzoni c. *La Divina Commedia*

_____ 4. Guareschi d. *Don Camillo*

Answers: 1-c, 2-a, 3-b, 4-d.

Comment:

Matching is a useful way to test specific literary knowledge.

4.3.3 Global Testing

Some teachers prefer items which do not focus on one specific learning problem; such items specifically involve more than one skill, language area and testing device at the same time. This technique has been called "global testing." The following is offered as a typical example:

Instructions:

You will hear an answer to a question. Match it with its written representation on your answer sheet and write a question which could have provoked the answer.

Spoken Answer: Non ne ho.

Answer Sheet: A. Non ne ha.
B. Non le ho.
C. Non ne ho.

Second Phase:

Student writes (for instance): Hai dei fiammiferi?, or any other question that could have preceded the answer.

Comment:

The skills being tested in this example are listening comprehension, reading, and writing. The language areas involve tense identification (in the listening and reading parts) and several problems in the writing part, such as the partitive construction, vocabulary and verb inflection. The testing devices are matching and transformation. In using such items the teacher must be prepared to judge the consequences of each possible error. If the student misses the first answer, his subsequent answers will also be incorrect. There must also be some well understood procedure for classifying possible errors if the teacher is to discern the student's weaknesses. Grading such items can be difficult if partial credit is given. Global testing may be used effectively, however, whenever the student has acquired a degree of moderate fluency and the teacher wishes to evaluate his versatility in the foreign language.

Chapter IV--RUSSIAN

PREPARING TEST ITEMS

4.1 INTRODUCTION

4.1.1 It is not easy to amass a large repertoire of effective test items. In addition to the difficult task of writing understandable instructions and credible distractors, the teacher must worry constantly about how the test fits what he has taught. In foreign-language work, it is often desirable for a test to sample all the language skills; on such a test, testing devices must be used which are appropriate to each skill. We hope that the teacher will find the following sections helpful in making his own test items. This manual claims neither novelty nor completeness. In Chapter IV, we have tried, first, to be explicit about the essential elements of successful foreign-language tests and then to organize them into a useful working outline accompanied by concrete examples.

4.2 THE ESSENTIAL ELEMENTS OF A TEST

4.2.1 There are three basic sets of elements in any language test: (1) the language skills, (2) specific testing devices, and (3) the particular language areas to be tested.

4.2.2 The Language Skills

4.2.2.1 Teachers generally recognize four fundamental skills: (1) listening comprehension, (2) speaking, (3) reading, and (4) writing. In audio-lingual programs, listening comprehension is considered to have primacy over the others. The position taken is that one cannot speak, read or write meaningfully in a language without being able to understand that language. The other three skills are also interrelated, but it is not clear that speaking is as important for reading and writing as listening comprehension is for all of them. Even in the native language, the reading skill is often--perhaps always--more highly developed than the ability to write.

4.2.2.2 Tests must take into consideration which specific skills are to be evaluated. When the reading skill has been the main concern of the course, for instance, tests should not emphasize listening comprehension; in audio-lingual programs, which generally begin with listening comprehension and speaking, reading and writing should not be tested in the early stages of instruction.

4.2.2.3 Of the four skills, listening comprehension and reading are easiest to evaluate objectively. The difficulties inherent in testing writing and speaking are readily apparent. With listening and reading it is a comparatively simple matter to control the student's responses, but it is difficult to do so when the student is to react orally or in writing. Judgment of spoken and written responses involves much personal interpretation by the teacher; it is difficult to restrict the student to only one correct response. The burden of grading and administering oral-response tests makes them unwieldy; although written answers are somewhat easier to grade, they are just as difficult to control as spoken responses. Having a number of teachers grade written responses may actually increase the degree of subjectivity, although some experts disagree.[1]

4.2.3 Testing Devices

4.2.3.1 The most commonly used devices in foreign-language testing are: (1) multiple-choice items, (2) completion items, (3) matching sets, (4) true/false items, (5) translation, (6) dictation, and items involving (7) expansion, (8) transformation, and (9) substitution. The last three are especially useful to test speaking and writing. Of the other devices, true/false items and translation are the ones most often debated. It is very difficult to phrase a true/false item so that it is clearly and unquestionably true or false but not transparently so. Translation has several shortcomings. Using translation from the native language into the foreign language to discover the student's control of specific problem areas (the use of certain cases with special verbs, for example, in Russian) is ineffective in cases where the student can find an alternate possibility (a synonymous verb taking a different case, in this instance). Although translation from the foreign language to the native language is easier to control, its effectiveness is also somewhat limited because the student can usually find clues to the meaning of the phrase or word to be translated in language features (context, word order, etc.) other than the specific feature being tested. Therefore, unless the teacher can control quite strictly the context of a particular word or phrase, it is difficult to determine the student's actual command of the matter being tested. Extensive use of translation is obviously subject to attack in any curriculum that consciously emphasizes direct use and knowledge of the foreign language with as little recourse as possible to the student's native language. Perhaps translation is more accurately a specialized skill than a testing device. There is certainly no proof that the ability to understand, speak, read, and write a foreign language also guarantees the ability to translate freely between that language and one's own language. Indeed, in institutions with a sound foreign-language program designed for specialists, translation is often taught in a separate course to which students are admitted only after they have demonstrated a command of the other skills.

---

[1] See, for example, The Measurement of Writing Ability by Fred Godshalk, Frances Swineford and William Coffman, College Entrance Examination Board, New York (1966). The articulation of norms for written composition differs from language to language. For an illuminating discussion of procedures, see How the French Boy Learns to Write by Rollo Walter Brown, Harvard University Press (1915), reprinted by the National Council of Teachers of English (1963).

4.2.3.2 *Multiple-choice items* (i.e., choice of one correct answer from two or more possible answers) and *matching sets* are easy to control since variation in answers is strictly limited. To be effective, multiple-choice items must have distractors which are sufficiently similar to the correct choice to attract the student who does not really know what the correct answer is. Incorrect answers can reveal the individual student's problems. A wrong choice between стол and столь or брат and брать in Russian, for example, may mean that the student needs more practice in distinguishing hard from soft consonants. By increasing the number of possible answers, the teacher can gain more information about the individual student's weaknesses and needs.

4.2.3.3 There are two ways of constructing *matching sets*: (1) the same number of items may be supplied in each of two columns, or (2) more items may appear in the column from which the "answers" are to be selected, thereby leaving a residue when the answers have been chosen. In either case, matching is a type of interlocking multiple choice. Guessing, however, can be a factor in both matching and multiple-choice items. A formula for penalizing random choices can be used; it should be explained to the students in an effort to discourage them from making irresponsible guesses.

4.2.3.4 *Expansion items* test the student's knowledge of the construction of sentences in the foreign language. The student may be asked to expand a sentence in a number of different ways. In the following example, he is being tested on his knowledge of possessive modifiers:

Instructions: You will hear a pair of short sentences. Add the appropriate possessive modifier to the second sentence, referring to the owner indicated in the first sentence.

Student hears: Вот книга Ивана. Иван забыл книгу.
Вот тетрадь Бориса. Иван возьмёт тетрадь.

Answers: Иван забыл свою книгу.
Иван возьмёт его книгу.

4.2.3.5 Transformation items are particularly suited to evaluating writing. The student is given instructions and a model; for example, he is given a sentence with a verb in the present tense to be changed to a past tense. A series of sentences follows with the verb in the present tense which the student is to change to the past. Transformation items can be written to test more complex matters such as making active sentences passive, inserting subordinate clauses into simple sentences to make complex sentences, etc. It is always advisable to furnish a model item as part of the instructions.

4.2.3.6 substitution items, the student is asked to replace a word, part of a word, or phrase with another;

Instructions:	Replace the underlined words with others which are equally correct.
Sentence:	Мы ещё не знаем всех студентов.
Possible Answer:	Мы ещё не знаем всех девушек.

4.2.3.7 Dictation is a well-known device in foreign-language instruction. However, it is difficult to grade straight dictation. The teacher must decide how to sort out and weight each student's individual errors. One way to avoid many problems is to give the student a partially filled-in answer sheet. As the teacher reads the dictation, the student writes in only what has been omitted from the text he has before him. The blanks may become more numerous as the student progresses in his ability to take dictation.

4.2.4 The Language Areas

4.2.4.1 The third basic set of elements involved in the preparation

3

of foreign-language tests, the language areas to be tested, is the least clearly defined and by far the largest of the three. Opinions about the relative importance of various language areas differ widely.

4.2.5 Contrastive Analysis

4.2.5.1 The student's native language is carefully contrasted with the foreign language he is studying because it is believed that the language learner will make numerous predictable errors when he incorrectly transfers the habits he has acquired in his native language to the language he is studying. In other words, the learner tends to express himself in the foreign language in terms of the rules of the language he already knows. He will automatically assume that whatever is not clearly different in the foreign language is the same as in his own language. A thorough knowledge of the differences that exist between the two languages can therefore help the teacher to anticipate problems.

4.2.5.2 A foreign-language test could well consist very largely of items on the areas where the two languages differ. If the teacher knows, for example, that Russian grammatical gender is different from English gender, he can expect gender to be a problem for the English-speaking learner and may therefore devote a large proportion of his testing to gender in the early stages of instruction.

4.2.5.3 Wherever the native language of the student has a feature which is lacking or simpler in the foreign language, the problem is relatively easy to solve. Where Russian has no articles, for example, English has a comparatively complicated system. As a result, the

English-speaking learner of Russian has to learn to express himself without articles while the Russian student of English has to learn to use correctly the English definite and indefinite articles.

4.2.5.4 Both the contrasts and the ways in which we state them depend on the theoretical premisses we employ in our contrastive analysis. We suggest the following outline as a good guide to the arrangement of basic language matters:

I. Phonology (Sound System)
  a. phonemes (distinctive classes of sounds)
  b. distribution of phonemes (including diphthongs and consonant clusters)
  c. intonation (pitch changes in the voice to indicate statements, questions and emotional coloring)

II. Morphology (Formation of Words)
  a. morphemes (the basic units of grammar)
  b. inflection (changes in words according to number, person, case, etc.)
  c. derivation of words (formation of new words by compounding or by the addition of prefixes and suffixes)

III. Syntax (Arrangements of Words in Larger Constructions)
  a. phrases
  b. clauses
  c. sentences

4.2.6 Vocabulary

The size and nature of a student's vocabulary is obviously important in language teaching. Various word counts exist,[2] but there is surprisingly little agreement about what words should be taught at each level of language instruction. However, since the teacher must test vocabulary, he should select vocabulary items as shrewdly as he can to test those items that seem most useful. For purposes of classification, idioms should be included in 'vocabulary' It is also often useful to group words into families or general categories.

4.2.7 Style

Russians make systematic use of a number of different styles of discourse of varying degrees of formality and impersonality, both orally and in writing. Some distinctions in style affect even basic grammatical forms. Such is the case with the forms of polite and familiar address in Russian. At least the most important distinctions of style are treated in most language texts and should be tested.

---

[2] An interesting study has been done for English at Brown University (see the forthcoming publication, Computational Analyses of Present Day American English, edited by Henry Kucera and W. N. Francis). In a text of 1,014,235 words, there were only 50,457 different words; 42,540 of these occurred only ten times or less.

4.2.8 Literature

Literature is generally considered an essential part of foreign-language instruction. In audio-lingual programs literature is not ignored; rather, it is not treated until students can read with some ease and with sensitivity to literary values. The choice of specific literary selections varies, of course, from teacher to teacher and from class to class.

4.2.9 Culture

There is no generally accepted, useful definition of culture in the context of foreign-language teaching. Culture is understood either in a narrow sense, meaning understanding of the literary and artistic values general among the educated speakers of a language, or it is extended to include the whole way of life and the history of a people, the geography of the land they live in and their political, religious and social institutions. Teachers seem to agree, however, that culture is an essential part of language instruction and that they should test their students on it.

4.2.10 A Brief List of Language Areas

The following list of language areas is offered:

(1) distinctive sound classes (phonology) (e.g., the stressed vowel sounds in там, едет, мёд, вижу, дом; hard vs. soft consonants, etc.);

(2) orthography (e.g., unstressed vowels);

(3) consonant clusters (e.g., щ , double consonants, etc.);

(4) statement intonation;

(5) question intonation;

(6) basic grammar units (e.g., cases, aspects);

(7) inflectional procedures (e.g., nominal inflection);

(8) derivation (e.g., formation of diminutives, etc.);

(9) phrasal syntax;

(10) clausal syntax;

(11) sentence syntax;

(12) vocabulary (including both single words and idioms);

(13) style (e.g., 'colloquial' vs. 'literary' Russian);

(14) literature (names of well-known writers, reading selections);

(15) culture (accepted ways of speaking in specified social situations, ways of addressing officials, etc.).

The list is open; additional features can be added to it and the ones given can be made more detailed.

4.3 WORKING OUTLINES FOR TEST WRITING

The skills and testing devices have been arranged on the following chart:

Chart No. 1: Skills and Testing Devices

Skills	Testing Devices											
	MC	CO	MA	T/F	TR	DIC	EX	T	SUB	RA	EC	R
Listening Comprehension	X		X		X							X
Speaking	X	X			X		X	X	X	X		X
Reading	X		X	X	X					X		
Writing	X	X			X	X	X	X	X		X	

Key:
MC - multiple choice
CO - completion
MA - matching
T/F - true/false
TR - translation
DIC - dictation
EX - expansion
T - transformation
SUB - substitution
RA - reading aloud
EC - extended composition
R - repetition
X - effective use

An examination of the chart leads to some general observations on the appropriateness of particular testing devices to specific skills:

(1) Generally speaking, completion items are hard to control. In order to assure objectivity, completion items must be restricted to absolutely unambiguous possibilities. (Visuals can be useful for this purpose).

(2) Ambiguity is a major problem in true/false questions. The teacher should be especially cautious in his use of ungrammatical

sentences which the student is to identify as "incorrect." Just as correct stimuli appear to reinforce learning, incorrect stimuli may create uncertainty where the student is well prepared.

(3) The use of translation to test all four skills is open to question. Problems arise in restricting translation so as to test only specific points. Translation can also be considered a distinct skill; as such, it, too, can be tested by a number of devices, such as multiple choice, completion, matching and true/false items.

(4) Expansion and transformation items are useful devices for testing speaking and writing, but they are inappropriate devices for testing the receptive skills of listening and reading.

It is reasonable to suppose that some testing devices are more suitable than others to pupils of different ages. In the absence of evidence from controlled experimentation, the individual teacher must rely on his own judgment.

4.3.1 Chart No. 2, combining language areas and skills, is a check list for the preparation of test items. By filling it with actual test items of different kinds, the teacher can achieve a broad coverage of language problems. The following samples illustrate item composition. Each item is numbered in cross reference to Chart No. 2.

Chart No. 2: Language Areas and Skills

Language Areas	Skills			
	Listening Comprehension	Speaking	Reading	Writing
Sound Classes (Phonemes)	(3)	(5)	(5)	
Orthography	(19)			(19)
Consonant Clusters		(5)		
Statement Intonation	(4)	(6)		
Question Intonation	(4)			
Grammar Units (Morphemes)			(12)	
Inflection	(8)(9)	(8)(9)	(8)(9)	(8)(9)
Derivation	(10)	(10)	(10)	(10)
Syntax: Phrases	(2)(6)	(6)	(2)(11)(12)	
Syntax: Clauses	(2)		(2)(13)(16)	(16)
Syntax: Sentences	(2)(7)(8)	(7)(8)	(2)(7) (8)(14)	(7)(8)(15)
Vocabulary	(1)(2)		(1)(2)(18)	(18)
Style			(17)	(17)

4.3.2 Sample Test Items

(1) Language Area: Vocabulary

Skill: Listening comprehension

Testing Device: Multiple choice

Instructions: You will hear several questions in Russian one time each. After each question you will hear four short responses to the question. These also will be spoken only once. After you have heard all four, put down the letter (a, b, c, d) of the response best suited to the question.

1. Как долго вы стояли в очереди?

   a. Вчера.
   b. Всё утро.
   c. В час.
   d. В ГУМе.

2. Где покупают продукты?

   a. На заводе.
   b. В гастрономе.
   c. В ресторане.
   d. В аптеке.

3. Хотите пойти на концерт?

   a. С удовольствием!
   b. Вот досада!
   c. Всего хорошего!
   d. Поздравляю вас!

Answers:
1. b
2. b
3. a

Comment: This item could also be used in written form to test reading.

(2) Language Areas: Vocabulary/Syntax

Skill: Listening comprehension

Testing Device: Multiple choice

Instructions: You will hear twice a descriptive passage of moderate difficulty. After you have heard the passage you will be asked five questions on the passage. Listen to the choices, then select from the choices the reply which is correct according to the information in the passage. Mark on your answer sheet the letter which corresponds to your choice.

Двести километров к северо-востоку от Москвы находится колхоз "Красный октябрь". Вокруг колхоза--сплошной лес, который состоит большей частью из елей и берёз. Лес, как и везде на севере России, даёт особенный характер жизни людей в этой части Земли.

Почва в этом районе не очень хороша, но колхозные поля--благодаря науке--дают довольно хорошие урожаи. В колхозе производятся рожь, пшеница, овёс и картошка. Частные огороды колхозников дают вдобавок овощи--капусту, лук, морковь и огурцы. Фруктовых деревьев нет, потому что зимы слишком суровые.

Дома и другие здания в колхозе все деревянные. Кроме частных домов есть правление колхоза, в одном здании начальная школа, детский сад и библиотека, и в отдельном здании, зал для собраний и танцев. Недалеко от колхоза есть маленькая станция, где можно сесть на поезд в Москву.

На колхозе "Красный октябрь живёт--если считать детей--полтораста человек. Хотя работают все взрослые люди, всегда есть недостаточно рабочих. Это объясняется очень просто: молодые люди не хотят оставаться в деревне, в результате чего колхоз теряет больше половины своей молодёжи.

Student hears: (#1) Сейчас слушайте вопросы. Вопросы повторяются только один раз. Вопрос номер один:

В какой части СССР находится колхоз "Красный октябрь"?

a. На юге.
b. На севере.
c. На востоке.

(#2) Вопрос номер два: Что даёт особенный характер жизни людей в колхозе?

a. Окружающие леса.
b. Близкие озёра.
c. Короткие лета.

(#3) Вопрос номер три: Почему нет фруктовых деревьев в колхозе?

a. Почва в этом районе такая плохая.
b. Зимой слишком холодно в колхозе.
c. Колхоз не назначен для производства фруктов.

(#4) Вопрос номер четыре: Если считать детей, сколько человек живёт в колхозе "Красный октябрь"?

a. Около двухсот.
b. Сто пятьдесят.
c. Меньше ста.

(#5) Вопрос номер пять: Почему колхоз теряет больше половины своей молодёжи?

a. В колхозе нет достаточно работы.
b. Жизнь лучше в других деревнях.
c. Молодые люди предпочитают жить в городе.

Comment: This is a difficult type of item. It should be used only when the student has acquired a rather substantial amount of Russian. Items of this sort can also be easily adapted to test reading.

(3) **Language Area:** Sound classes (phonology)

Skill: Listening comprehension

Testing Device: True/False

Instructions: You will hear a number of Russian words. Listen especially to the first consonant. Indicate whether it was hard or soft by writing h for hard, s for soft on your answer sheet. Here is an example:

Example: дядя

You would write s on your answer sheet.

Student hears:
1. аабыл
2. век
3. лягу
4. дом
5. тьма

Answers:
1. h
2. s
3. s
4. h
5. s

Comment: Items of this kind are best considered modified true/false items. The student is asked to write h meaning "it is true, that the consonant is hard" or s meaning "it is false that the consonant is hard; it is soft."

(4) Language Area: Intonation

Skill: Listening comprehension

Testing Device: Multiple choice

Instructions: You will hear a number of utterances in Russian. Included are neutral statements, emphatic statements, and questions. As you hear each, indicate on your answer sheet the type of utterance it is. Use N for neutral statement, E for emphatic statement, and Q for question.

Student hears:
1. Я давно вас не видел!
2. Вы читали эту книгу?
3. Борис может это купить?
4. Мария опять ссорилась с подругой!
5. Вчера был четверг.

Answers:
1. E
2. Q
3. Q
4. E
5. N

Comment: The Russian utterances can be presented quite rapidly, since the student is listening for distinctive melodies more than for lexical or grammatical features. By moving at a fairly rapid pace the student is prevented from cross-comparing his answers, so that the teacher can be reasonably certain that he has answered on the basis of his first impression. If a large number of utterances are presented, it is always clear whether the student recognizes the contours. If he does not recognize some or all of the contours, a large number of responses makes clear precisely where the difficulty lies.

(5) Language Area: Sound classes

Skills: Listening comprehension/Speaking

Testing Device: Repetition

Instructions: You will hear several sentences. Repeat each sentence, imitating the speaker's pronunciation as closely as you can.

Student hears:

1. Саша живёт здесь уже шесть месяцев.
2. На стене висело шесть старых картин.
3. Все чехи играют в шахматы.

Answers: The teacher listens for the following:

1. ш, ж
2. ст
3. х

Comment: This same item could be presented in written form. It would then test two abilities: the student's ability to interpret Russian orthography correctly and his ability to pronounce the language acceptably.

(6) Language Area: Syntax (phrases)

Skills: Speaking/Listening comprehension

Testing Device: Expansion

Instructions: You will be given several simple Russian sentences, one at a time. After you hear each sentence, you may repeat it aloud once to help you remember it. Next you will be given an adjective. Insert the correct form of the adjective into its proper place in the sentence. You will then be given a Russian number from one to five. Insert the correct forms of both the adjective and the number into their proper places in the original sentence. Remember to make all necessary changes.

Student is told: (1) Я купил рубашку.

Student is told: новый

Student says: Я купил новую рубашку.

Student is told: три

Student says: Я купил три новых рубашки.

Student is told: (2) Дайте им стакан воды.

Student is told: большой

Student says: Дайте им большой стакан воды.

Student is told: два

Student says: Дайте им два больших стакана воды.

Student is told: (3) Мы видели девушку.

Student is told: красивый

Student says: Мы видели красивую девушку.

Student is told: четыре

Student says: Мы видели четырёх красивых девушек.

Comment: This type of item can be recorded so that the teacher can play the students' responses back a second time to evaluate their pronunciation, intonation, fluency, etc.

(7) Language Area: Syntax (sentences)

Skills: Reading/Writing, or:
Listening comprehension/Speaking

Testing Device: Transformation

Instructions: You will be given several sentences. You are to change each sentence so that the present subject will appear as an object, and the present object, as subject in a new sentence. Here is an example:

Example: Он без меня не пойдёт.
Answer: Я без него не пойду.

1. Алексеев вчера был на лекции с Ниной.

2. Она это сделала для нас.

Answers:
1. Нина вчера была на лекции с Алексеевым.
2. Мы это сделаем для неё.

Comment: Items of this type can be presented equally effectively orally or in written form. When they are presented orally they should be administered so that the student's responses are recorded to enable the teacher to evaluate them with care.

(8) Language Area: Syntax (sentences)/Inflection

Skills: Speaking/Listening comprehension

Testing Device: Transformation

Instructions: You will be given several Russian sentences. Each sentence contains only singular words or only plural words. You are to make all singular words plural, and all plural words singular. Here is an example:

Example: Студентки читали дома.
Answer: Студентка читала дома.

Student hears: (#1) Эти люди ходят к хорошим докторам.

Student says: Этот человек ходит к хорошему доктору.

(#2) Чей этот ключ?

Чьи это ключи?

(#3) Кроме этого англичанина, в нашей школе нет иностранца.
Кроме этих англичан, в наших школах нет иностранцев.

Comment: Items of this type can also be used to test writing (and reading, incidentally).

(9) Language Area: Inflection

Skills: Speaking/Listening comprehension

Testing Device: Transformation

Instructions: You will be given several sentences in which the verb forms are all of the past tense. Repeat the sentences changing the verbs to the present tense.

Example: Как вы это делали?
Answer: Как вы это делаете?

1. Отец вёл сынка в парк.

2. Почему только девушки танцевали?

3. Эти люди искали комнату.

Answers:
1. ведёт
2. танцуют
3. ищут

Comment: Items of this type can also be administered in writing to test writing (and reading, incidentally).

(10) Language Area: Derivation

Skills: Speaking/Listening comprehension

Testing Device: Transformation

Instructions: You will hear a number of statements containing diminutives of names. Repeat each statement substituting the full form of the name.

Example: Саша сейчас болен.
Answer: Александр сейчас болен.

Student hears:
1. У Оли нет коньков.

2. Маша теперь в Москве.

3. Мы с Вовой пошли на выставку.

4. Я позвоню Грише сегодня вечером.

Answers:
1. У Ольги нет коньков.
2. Мария теперь в Москве.
3. Мы с Владимиром пошли на выставку.
4. Я позвоню Григорию сегодня вечером.

Comment: It seems best to stretch the notion of transformation to include phenomena of this sort. Items of this kind can also be administered in written form to test writing (and reading).

(11) Language Area: Syntax (phrases)

Skill: Reading

Testing Device: True/False

Instructions: Indicate whether the two sentences have the same meaning by checking true or false.

Моя книга в шкафу.

Книга у меня в шкафу.

( ) True
(x) False

Answer: As indicated.

Comment: If the student answers this item on the basis of understanding (rather than guessing), it indicates that he is aware that у-constructions are prepositional phrases with adverbial force, structurally incapable of modifying nouns. The meaning of both sentences is close, to be sure, but the first person is present in the subject in the first sentence, in the predicate in the second, and there is no meaning of possession whatsoever in the second sentence.

(12) Language Area: Grammar units

Skill: Reading

Testing Device: Multiple choice

Instructions: Place an X in the parentheses before the preposition which correctly completes each of the following sentences.

1. Мальчики идут домой _______ почты.

( ) из
( ) от
(x) с
( ) у

2. Собака убежала ______ угол, я не видела, куда она спряталась.

( ) от
(x) за
( ) до
( ) про

Comment: This item can also be considered to test knowledge of phrase syntax.

(13) Language Area: Syntax (clauses)

Skill: Reading

Testing Device: Multiple choice

Instructions: Place an X between the parentheses before the form of the relative adjective which correctly completes each of the sentences.

1. Вазу, _______ стоит на столе, мне подарила мать.

( ) которую
( ) которой
(x) которая
( ) которому

2. Это люди, ______ нельзя верить.

( ) которые
( ) которых
(x) которым
( ) которыми

Answers: As indicated.

(14) Language Area: Syntax (sentences)

Skill: Reading

Testing Device: Multiple choice

Instructions: Place an X in the parentheses before the verbs which correctly complete each of the following Russian sentences.

1. Нам пообещали, что _______ к нам сегодня вечером по дороге домой.

( ) зашли
( ) зашли бы
(x) зайдут
( ) заходят

2. Вчера ночью им нельзя ______ спать.

( ) был
( ) была
(x) было
( ) были

Answers: As indicated.

(15) Language Area: Syntax (sentences)

Skill: Writing

Testing Device: Transformation/Expansion

Instructions: Given below are several groups of words. Write one sentence for each group. All the words in the group should appear in the sentence, but you may add such others as are required to complete the sense of your sentence.

1. жить, около, город
2. пойти, лес, гриб
3. хотеть, подарок, брат

Possible Answers:

1. a. Катя живёт в городе, около вокзала.
   b. Они жили около города.
2. a. Давайте пойдём в лес за грибами.
   b. Молодые люди пошли в лес за грибами.
3. a. Кто хочет купить подарок для брата?
   b. Мой брат хотел купить подарок в ГУМе.

Comment: Many correct sentences can be written using the words supplied. The problem is essentially global, since the student is permitted to add words. A variation on this type is to require that the responses be connected sequentially, so that the end result is a dialogue or a paragraph rather than separate, unrelated sentences.

(16) Language Area: Syntax (clauses)

Skills: Writing/Reading

Testing Device: Transformation/Expansion

Instructions: Given below are several sentences containing adverbial participles ("gerunds"). These have been underlined. Rewrite each underlined participle as an adverbial clause.

Example: Познакомившись, молодые люди стали хорошими друзьями.

Когда познакомились ________________
Как только познакомились ______________

1. Сняв телефонную трубку, Алексей Иванович тихо сказал "алло".
2. Не найдя работы в Нью-Йорке, молодые люди решили поехать в Калифорнию.
3. Завтракая, он слушал новости по радио.

Answers:

1. Когда снял ___________
   Как только снял ______
2. Потому, что не нашли __________
   Так как не нашли ______________
3. Пока он завтракал _______________________
   В то время, как завтракал ______________

Comment: An item of this type, where several stylistic variants are possible in the answer, can be used only with sophisticated students.

(17) Language Area: Style

Skills: Writing/Reading

Testing Device: Extended composition

Instructions: Given below is a short descriptive passage in Russian. Using this passage as a basis, write a dialogue of at least four lines. Some material may be added, and it is not necessary to include in the dialogue everything that appears in the passage.

Мой отец механик. Он работает в мастерской. С ним работают и другие механики. В мастерской починяют автомобили и грузовики. Иногда отец мне позволяет ходить в мастерскую. Я очень люблю наблюдать, как там пользуются инструментами. Когда я в мастерской, я всегда слушаюсь механиков, потому что там опасно.

Suggested Answer:

--Где работает твой отец?
--Мой отец механик. Он работает в мастерской. Там починяют автомобили и грузовики.
--Твой отец тебе позволяет ходить в мастерскую?
--Да, иногда. Я очень люблю наблюдать, как там пользуются инструментами. Но надо слушаться механиков, потому что в мастерской опасно.

Comment: This type of item cannot be used until the student has done some reading and is aware of the differences between spoken colloquial style and simple written style. The restatement can be in either direction (i.e., with either dialogue or expository passage as cue).

(18) Language Area: Vocabulary

Skills: Writing/Reading

Testing Device: Completion

Instructions: Fill in the blanks in the following sentences with words suggested by the sentence context.

1. Это случилось в 1125-ом году, то есть, в ____________________ веке.
2. В ГУМе турист заплатил два __________, восемь ____________ за подарок.
3. ______ добры, идите туда с нами!
4. Зимой холодно, а летом __________.

Answers:

1. двенадцатом
2. рубля, копеек
3. Будьте
4. жарко

Comment: This device is very effective for testing vocabulary. No recourse to English is necessary. Moreover, it follows the principle that vocabulary cannot be divorced from context. Items of this kind are not easy to write, however, for the context which is invented must be absolutely unambiguous. A further weakness of this device is that a student's general knowledge also is often tested. For example, to answer the first question above, he must not only know the Russian ordinal numerals, but also the convention followed in Russian for numbering the centuries. Questions of this kind are best adapted to testing closed semantic subsystems in the language, such as parts of the body, kinship terms, telling time, names of the days of the week and of the months, etc. Ambiguity is least likely to occur in these areas.

(19) Language Area: Orthography

Skills: Writing/Listening

Testing Device: Dictation

Instructions: You have been given a Russian text containing blanks. As this passage is read to you, fill in the missing word or words which should appear in the blanks.

Complete text (to be read to the student):

В детстве, в первую пору молодости, человек учиться в школах: уроки учителей имеют ту цель, чтобы сделать юношу образованным человеком. Но когда он выходит из школы, перестаёт учиться, его образование поддерживается чтением, то-есть вместо прежних учителей, которых слушал мальчик и юноша, взрослый человек имеет одну учительницу - литературу.

/Чернышевский/

Text with blanks:

В детстве, ________ пору ________, человек учиться в
школах: уроки ________ имеют ту _____, чтобы ______ юношу
_______________человеком. Но когда он выходит ________, _____
________ ________, его образование _________________ чтением,
________ вместо прежних учителей, ___________ слушал мальчик и
юноша, взрослый человек ________ одну учтельницу - литературу.

Comment: Omitted are words (1) which contain sounds or sound combinations difficult for English-speaking persons, or (2) whose spelling diverges rather far from the spoken form of the word. Also omitted is the word **имеет**. This word is actually not difficult to spell, but some students have difficulty reconciling themselves to writing two consecutive e's. Listening comprehension is tested in items of this kind, in addition to spelling.

Chapter IV--SPANISH

PREPARING TEST ITEMS

4.1 INTRODUCTION

4.1.1 It is not easy to amass a large repertoire of effective test items. In addition to the difficult task of writing understandable instructions and credible distractors, the teacher must worry constantly about how the test fits what he has taught. In foreign-language work, it is often desirable for a test to sample all the language skills; on such a test, testing devices must be used which are appropriate to each skill. We hope that the teacher will find the following sections helpful in making his own test items. This manual claims neither novelty nor completeness. In Chapter IV, we have tried, first, to be explicit about the essential elements of successful foreign-language tests and then to organize them into a useful working outline accompanied by concrete examples.

4.2 THE ESSENTIAL ELEMENTS OF A TEST

4.2.1 There are three basic sets of elements in any language test: (1) the language skills, (2) specific testing devices, and (3) the particular language areas to be tested.

4.2.2 The Language Skills

4.2.2.1 Teachers generally recognize four fundamental skills: (1) listening comprehension, (2) speaking, (3) reading, and (4) writing. In audio-lingual programs, listening comprehension is considered to have primacy over the others. The position taken is that one cannot speak, read or write meaningfully in a language without being able to understand that language. The other three skills are also interrelated, but it is not clear that speaking is as important for reading and writing as listening comprehension is for all of them. Even in the native language, the reading skill is often--perhaps always--more highly developed than the ability to write.

4.2.2.2 Tests must take into consideration which specific skills are to be evaluated. When the reading skill has been the main concern of the course, for instance, tests should not emphasize listening comprehension; in audio-lingual programs, which generally begin with listening comprehension and speaking, reading and writing should not be tested in the early stages of instruction.

4.2.2.3 Of the four skills, listening comprehension and reading are easiest to evaluate objectively. The difficulties inherent in testing writing and speaking are readily apparent. With listening and reading it is a comparatively simple matter to control the student's responses, but it is difficult to do so when the student is to react orally or in writing. Judgment of spoken and written responses involves much personal interpretation by the teacher; it is difficult to restrict the student to only one correct response. The burden of grading and administering oral-response tests makes them unwieldy; although written answers are somewhat easier to grade, they are just as difficult to control as spoken responses.

Having a number of teachers grade written responses may actually increase the degree of subjectivity, although some experts disagree.[1]

4.2.3 Testing Devices

4.2.3.1 The most commonly used devices in foreign-language testing are: (1) multiple-choice items, (2) completion items, (3) matching sets, (4) true/false items, (5) translation, (6) dictation, and items involving (7) expansion, (8) transformation, and (9) substitution. The last three are especially useful to test speaking and writing. Of the other devices, *true/false items and translation* are the ones most often debated. It is very difficult to phrase a true/false item so that it is clearly and unquestionably true or false but not transparently so. Translation has several shortcomings. Using translation from the native language into the foreign language to discover the student's control of specific problem areas (the subjunctive, for example, in Spanish) is ineffective in cases where the student can find an alternate possibility (the infinitive, for example). Although translation from the foreign language to the native language is easier to control, its effectiveness is also somewhat limited because the student can usually find clues to the meaning of the phrase or word to be translated in language features (context, word order, etc.) other than the specific feature being tested. Therefore, unless the teacher can control quite strictly the context of a particular word or phrase, it is difficult to determine the student's actual command of the matter being tested. Extensive use of translation is obviously subject to attack in any curriculum that consciously emphasizes direct use and knowledge of the foreign language with as little recourse as possible to the student's native language. Perhaps translation is more accurately a specialized skill than a testing device. There is certainly no proof that the ability to understand, speak, read, and write a foreign language also guarantees the ability to translate freely between that language and one's own language. Indeed, in institutions with a sound foreign-language program designed for specialists, translation is often taught in a separate course to which students are admitted only after they have demonstrated a command of the other skills.

---

[1] See, for example, *The Measurement of Writing Ability* by Fred Godshalk, Frances Swineford and William Coffman, College Entrance Examination Board, New York (1966). The articulation of norms for written composition differs from language to language. For an illuminating discussion of procedures, see *How the French Boy Learns to Write* by Rollo Walter Brown, Harvard University Press (1915), reprinted by the National Council of Teachers of English (1963).

4.2.3.2 Multiple-choice items (i.e., choice of one correct answer from two or more possible answers) and matching sets are easy to control since variation in answers is strictly limited. To be effective, multiple-choice items must have distractors which are sufficiently similar to the correct choice to attract the student who does not really know what the correct answer is. Incorrect answers can reveal the individual student's problems. A wrong choice between pero and perro or caro and carro in Spanish, for example, may mean that the student needs more practice in distinguishing /r/ from /R/. By increasing the number of possible answers, the teacher can gain more information about the individual student's weaknesses and needs.

4.2.3.4 There are two ways of constructing *matching sets*: (1) the same number of items may be supplied in each of two columns, or (2) more items may appear in the column from which the 'answers' are to be selected, thereby leaving a residue when the answers have been chosen. In either case, matching is a type of interlocking multiple choice. Guessing, however, can be a factor in both matching and multiple-choice items. A formula for penalizing random choices can be used; it should be explained to the students in an effort to discourage them from making irresponsible guesses.

4.2.3.5 *Expansion items* test the student's knowledge of the construction of sentences in the foreign language. The student may be asked to expand a sentence in a number of different ways. In the following example, he is being tested on his knowledge of the positioning of adjectival modifiers in noun phrases:

Instructions: In each of the following sentences insert the word you are given:

Sentence No. 1: Ese señor es el embajador.

Insert: mexicano

Answer: Ese señor es el embajador mexicano.

Sentence No. 2: Ayer recibimos las noticias.

Insert: malas

Answer: Ayer recibimos las malas noticias.

4.2.3.6 Transformation items are particularly suited to evaluating writing. The student is given instructions and a model; for example, he is given a sentence with a verb in the present tense to be changed to a past tense. A series of sentences follows with the verb in the present tense which the student is to change to the past. Transformation items can be written to test more complex matters such as making active sentences passive, inserting subordinate clauses into simple sentences to make complex sentences, etc. It is always advisable to furnish a model item as part of the instructions.

4.2.3.7 In substitution items, the student is asked to replace a word, part of a word, or phrase with another:

Instructions:	Replace the underlined words with others which are equally correct.
Sentence:	Me lavé las manos.
Possible Answer:	Me lavé la cara.

4.2.3.8 Dictation is a well-known device in foreign-language instruction. However, it is difficult to grade straight dictation. The teacher must decide how to sort out and weight each student's individual errors. One way to avoid many problems is to give the student a partially filled-in answer sheet. As the teacher reads the dictation, the student writes in only what has been omitted from the text he has before him. The blanks may become more numerous as the student progresses in his ability to take dictation. Dictation has, however, only limited utility in teaching Spanish.

4.2.4 The Language Areas

4.2.4.1 The third basic set of elements involved in the preparation of foreign-language tests, the language areas to be tested, is the least clearly defined and by far the largest of the three. Opinions about the relative importance of various language areas differ widely.

4.2.5 Contrastive Analysis

4.2.5.1 The student's native language is carefully contrasted with the foreign language he is studying[2] because it is believed that the language learner will make numerous predictable errors when he incorrectly transfers the habits he has acquired in his native language to the language he is studying. In other words, the learner tends to express himself in the foreign language in terms of the rules of the language he already knows. He will automatically assume that whatever is not clearly different in the foreign language is the same as in his own language. A thorough knowledge of the differences that exist between the two languages can therefore help the teacher to anticipate problems.

4.2.5.2 A foreign-language test could well consist largely of items on the areas where the two languages differ. If the teacher knows, for example, that Spanish grammatical gender is different from English gender, he can expect gender to be a problem for the English-speaking learner and may therefore devote a large proportion of his testing to gender in the early stages of instruction.

---

[2] Cf. the Contrastive Structure Series (Spanish-English) published by the University of Chicago Press, Charles A. Ferguson, General Editor.

4.2.5.3 Wherever the native language of the student has a feature which is lacking or simpler in the foreign language, the problem is relatively easy to solve. Where Spanish has ¿(no es) verdad?, for example, English has a far more complicated set of expressions. As a result, the English-speaking learner of Spanish has to learn one invariable expression while the Spanish-speaking student of English has to learn to use correctly the set of English expressions aren't we, couldn't we, etc.

4.2.5.4 Both the contrasts and the ways in which we state them depend on the theoretical premises we employ in our contrastive analysis. We suggest the following outline as a good guide to the arrangement of basic language matters:

I. Phonology (Sound System)
    a. phonemes (distinctive classes of sounds)
    b. distribution of phonemes (including diphthongs and consonant clusters)
    c. intonation (pitch changes in the voice to indicate statements, questions and emotional coloring)

II. Morphology (Formation of Words)
    a. morphemes (the basic units of grammar)
    b. inflection (changes in words according to number, person, case, etc.)
    c. derivation of words (formation of new words by compounding or by the addition of prefixes and suffixes)

III. Syntax (Arrangements of Words in Larger Constructions)
   a. phrases
   b. clauses
   c. sentences

4.2.6 Vocabulary

The size and nature of a student's vocabulary is obviously important in language teaching. Various word counts exist,[3] but there is surprisingly little agreement about what words should be taught at each level of language instruction. However, since the teacher must test vocabulary, he should select vocabulary items as shrewdly as he can to test those items that seem most useful. For purposes of classification, idioms should be included in 'vocabulary'. It is also often useful to group words into families or general categories.

4.2.7 Style

Speakers of Spanish make systematic use of a number of different styles of discourse of varying degrees of formality and impersonality, both orally and in writing. Some distinctions in style affect even basic grammatical forms. Such is the case with the forms of polite and familiar address in Spanish. At least the most important distinctions of style are treated in most language texts and should be tested. For example, the student should be taught to recognize the stylistic incongruity of a combination like ¡*Oye, profesor*! as well as he does in its English equivalent *Hey, professor*!

---

[3] An interesting study has been done for English at Brown University (see the forthcoming publication, Computational Analyses of Present Day American English, edited by Henry Kucera and W. N. Francis). In a text of 1,014,235 words, there were only 50,457 different words; 42,540 of these occurred only ten times or less.

4.2.8 Literature

Literature is generally considered an essential part of foreign-language instruction. In audio-lingual programs literature is not ignored; rather, it is not treated until students can read with some ease and with sensitivity to literary values. The choice of specific literary selections varies, of course, from teacher to teacher and from class to class.

4.2.9 Culture

There is no generally accepted, useful definition of culture in the context of foreign-language teaching. Culture is understood either in a narrow sense, meaning understanding of the literary and artistic values general among the educated speakers of a language, or it is extended to include the whole way of life and the history of a people, the geography of the land they live in and their political, religious and social institutions. Teachers seem to agree, however, that culture is an essential part of language instruction and that they should test their students on it.

4.2.10 A Brief List of Language Areas

The following list of language areas is offered:

(1) distinctive sound classes (phonology) (e.g., the vowel sounds in p*a*so, p*e*so, p*i*so, p*o*so, p*u*so);

(2) diphthongs and longer combinations of vowels (e.g., s*ei*s, s*ie*te, f*ui*ste, estud*iái*s);

(3) consonants (e.g., the positional variants of /b/, /d/, /g/);

(4) statement intonation;

(5) question intonation (e.g., ¿*Va a Madrid*? versus *Va a Madrid*.);

(6) basic grammar units (e.g., *por/para*, *ser/estar*);

(7) inflectional procedures (e.g., verbal inflection);

(8) derivation (e.g., formation of diminutives, etc.);

(9) phrasal syntax (e.g., the positioning of adjectives: *el pobre hombre*, *el hombre pobre*);

(10) clausal syntax (e.g., use of noun clauses introduced by *que*, etc.);

(11) sentence syntax (e.g., sequence of tenses);

(12) vocabulary (including both single words and idioms like *hambre*, *tener hambre*);

(13) style;

(14) literature (names of well-known writers, reading selections);

(15) culture (accepted ways of speaking in specified social situations, ways of addressing officials, etc.).

The list is open; additional features can be added to it and the ones given can be made more detailed.

4.3 WORKING OUTLINES FOR TEST WRITING

The skills and testing devices have been arranged on the following chart:

Chart No. 1: Skills and Testing Devices

Skills	Testing Devices										
	MC	CO	MA	T/F	TR	DIC	EX	T	SUB	RA	EC
Listening Comprehension	X		X		X						
Speaking	X	X			X		X	X	X	X	
Reading	X		X	X	X					X	
Writing	X	X			X	X	X	X	X		X

Key:
MC - multiple choice
CO - completion
MA - matching
T/F - true/false
TR - translation
DIC - dictation
EX - expansion
T - transformation
SUB - substitution
RA - reading aloud
EC - extended composition
X - effective use

An examination of the chart leads to some general observations on the appropriateness of particular testing devices to specific skills:

(1) Generally speaking, completion items are hard to control. In order to assure objectivity, completion items must be restricted to absolutely unambiguous possibilities. (Visuals can be useful for this purpose).

(2) Ambiguity is a major problem in true/false questions. The teacher should be especially cautious in his use of ungrammatical sentences which the student is to identify as "incorrect." Just as correct stimuli appear to reinforce learning, incorrect stimuli may

create uncertainty where the student is well prepared.

(3) The use of translation to test all four skills is open to question. Problems arise in restricting translation so as to test only specific points. Translation can also be considered a distinct skill; as such, it, too, can be tested by a number of devices, such as multiple-choice, completion, matching and true/false items.

(4) Expansion and transformation items are useful devices for testing speaking and writing, but they are inappropriate devices for testing the receptive skills of listening and reading.

It is reasonable to suppose that some testing devices are more suitable than others to pupils of different ages. In the absence of evidence from controlled experimentation, the individual teacher must rely on his own judgment.

4.3.1 Chart No. 2, combining language areas and skills, is a check list for the preparation of test items. By filling it with actual test items of different kinds, the teacher can achieve a broad coverage of language problems. The following samples illustrate item composition. Each item is numbered in cross reference to Chart No. 2.

Chart No. 2: Language Areas and Skills

Language Areas	Skills			
	Listening Comprehension	Speaking	Reading	Writing
Sound Classes (Phonemes)	(1)			
Diphthongs	(7)			(7)
Orthography	(5)			(5)
Statement Intonation	(6)			
Question Intonation	(6)	(16)		
Grammar Units (Morphemes)	(2)	(16)	(2)(9)	
Inflection			(8)	(8)
Derivation			(10)	(10)
Syntax: phrases			(3)	
Syntax: clauses	(11)	(11)(16)	(3)(11)	(11)
Syntax: sentences	(11)	(11)(16)	(4)(11)	(11)
Vocabulary			(4)	(12)
Style			(13)	(13)
Literature			(14)	
Culture			(15)	

4.3.2 Sample Test Items

(1) *Language Area*: Sound classes (Phonemic discrimination)

Skill: Listening comprehension

Testing Device: Multiple choice

Instructions: In each of the following items, you will hear three Spanish words (*a*, *b*, and *c*); two of the three are the same, one is *different*. On your answer sheet circle the *letter* that corresponds to the one that is *different*.

Student hears: 1. pala - pala - palo

2. peso - beso - peso

Student sees: 1. a b c
2. a b c

Answers: 1. c
2. b

Comment:

In the first of these items, the student must distinguish vowels accurately in a position in which his native-language habits tend to mislead him. In the second item, he must hear a distinction in voicing without the additional distinction in aspiration between /p/ and /b/ to which he is accustomed in English.

(2) *Language Area*: Basic grammar (morphemes)

Skills: Listening comprehension/Reading

Testing Device: Multiple choice

Instructions: You will hear a statement or a question. From among the choices you are given, pick the *one* which could follow it most easily in a normal conversation.

Student hears: 1. ¿Te gusta el vino?

Student sees:
(A) Sí, el vino te gusta.
(B) No, no vino.
(C) Sí, pero no quiero ahora.
(D) No, no quiero ver el pino.

Student hears: 2. ¿Dónde vas a pasar el verano?

Student sees:
(A) Se paseaba por el patio.
(B) Se van con usted.
(C) No pasa nada.
(D) Vuelvo al sur.

Answers:
1. (C)
2. (D)

Comment:

The student must first understand accurately what he hears. He must then read accurately the four choices and select the one whose semantic and structural properties are such that it could follow the initial statement or question in a natural conversation.

(3) Language Area: Syntax (Phrases and Clauses)

Skill: Reading

Testing Device: Multiple choice

Instructions: Complete the following paragraph by inserting words from those listed in the column to the right. There are more words in the list than you will need. Do not use any word more than once.

Su fama no depende del todo ___(1)___ su participación en ___(2)___ política, ___(3)___ también ___(4)___ renombre ___(5)___ ha alcanzado en el mundo de ___(6)___ letras. ___(7)___ de ejercer la profesión de abogado en su ciudad natal, ___(8)___ trasladó a la capital ___(9)___ ha vivido ___(10)___ la fecha.

de
del
después
donde
dónde
en
hacia
hasta
la
las
pero
que
se
sino

Answers:

(1) de
(2) la
(3) sino
(4) del
(5) que
(6) las
(7) después
(8) se
(9) donde
(10) hasta

Comment:

An item of this kind tests simultaneously a large number of different syntactic matters. It tests the student's ability to handle the language easily and accurately in the context of a connected discourse.

(4) Language Areas: Vocabulary/Syntax

Skill: Reading

Testing Device: Multiple choice

Instructions: Read the following passage. After it you will find some incomplete sentences. For each sentence there are four suggested completions. Check the most appropriate completion. Base your choice on the information given in the passage.

Student sees:

Elisa continuó su camino a lo largo del río. De pronto se sintió fatigada, y sentándose en un tronco trató de olvidarse de su miedo contemplando la corriente. No supo cuánto tiempo permaneció sentada en el tronco. Un ruido de algo que se movía la hizo volver en sí, y vio con terror un caimán que salía de la maleza y se lanzaba al agua.

Elisa no pudo gritar, pero con un supremo esfuerzo se levantó y echó a correr. El miedo la había enloquecido y la impulsaba a actuar a ciegas. Una raíz la hizo caer, y cuando quiso levantarse sintió en el tobillo un dolor intenso.

(1) Elisa se detuvo para
___ contemplar la corriente.
_x_ descansar.
___ no cansarse.
___ pensar en sus temores.

(2) Elisa estuvo sentada a la orilla del río
___ mucho tiempo.
___ hasta que olvidó su miedo.
_x_ hasta que oyó algo.
___ sin volver en sí.

(3) Al huir del sitio, Elisa
_x_ tropezó.
___ hizo ruido.
___ salió de la maleza.
___ se fijó en el caimán.

(4) El resultado de su precipitada fuga fue que
___ se volvió loca.
___ se quedó ciega.
___ perdió el miedo.
_x_ se hizo daño.

(The correct choices have been indicated by an _x_.)

Comment:

Although the questions are framed to ascertain whether the student has really understood the lexical meaning of the words and expressions used in the passage, syntax is inevitably involved as well.

(5) Language Area: Orthography

Skills: Listening comprehension/Writing

Testing Device: Dictation

Instructions: You will hear a passage of connected prose read through once. It will then be dictated to you in short units. It will then be read through slowly to give you a chance to check your work.

Student hears (and writes):

No conozco a Roberto Morro Arangueren, autor de la obra sobre la civilización española que vamos a estudiar en este curso. El profesor se decidió y escogió ese libro por varias razones--las cuales francamente me parecen bastante vagas e incoherentes.

Comment:

The student must write /R/ correctly according to context (with a single r in initial, preconsonantal, postconsonantal or final position, rr in intervocalic position); he must use c and z correctly if the dictation distinguishes /θ/ and /s/, or he must use c, z and s correctly if the distinction is not made; he must use b and v accurately to represent /b/; and he must spell accurately a number of words (e.g., incoherentes) in which he must know whether or not to write an h. Spanish orthography is simpler than French, for instance, but it cannot be taken for granted The teacher must decide what specific points he wants to test and work out his grading accordingly.

(6) Language Area: Intonation

Skill: Listening comprehension

Testing Device: True/False

Instructions: You will hear a number of Spanish utterances. On your answer sheet circle S if the utterance is a statement and Q if it is a question.

Student hears:
1. No le gustan.
2. ¿Son hermanas?
3. ¿Van con su mamá?
4. Tú eres amigo de él.
5. Está en la cocina.

Student sees:
1. S Q
2. S Q
3. S Q
4. S Q
5. S Q

Answers:
1. S
2. Q
3. Q
4. S
5. S

Comment:

Notice that this item is best classified as a modified true/false item: S means 'it is true that it is a statement' and Q means 'it is false that it is a statement; it is a question.' In constructing items of this kind it is obviously necessary to use only utterances whose grammatical form itself is ambiguous; any utterance containing interrogative words (e.g., ¿Por qué no le gustan?) presents strong clues that make it impossible to test the student's perception of intonation alone.

(7) Language Area: Diphthongs, etc.

Skills: Listening comprehension/Writing

Testing Device: Dictation

Instructions: Write the words you will hear; each word will be read twice.

Student hears (and writes):

1. daos
2. peina
3. duda
4. vente
5. pena
6. das
7. deuda
8. veinte

Comment:

In order to test the accuracy of the student's perception of diphthongs, the teacher must choose pairs of words with a constant stress pattern; otherwise, the student is given irrelevant clues; he may well know, for instance, that in verbs like cerrar the diphthong ie occurs only in stressed syllables. In grading an item of this kind, the teacher must be prepared to make some sort of systematic allowance for errors other than those involving the vowel(s) of the stressed syllables: for example, a misspelling like bente in an item testing diphthongs should not count against the student as much as an error in the area actually being tested.

(8) Language Area: Inflection

Skills: Reading/Writing

Testing Device: Completion

Instructions: Fill in the blanks in the following text with words that are appropriate in this context in meaning and in form.

Student sees:

Alberto y su hermano Pablo se encuentran por la calle con su amigo Luis. Pablo tiene catarro y no oye bien.

Alberto: Buenos día, Luis. ¿Adónde vas?

Luis: Voy al centro. Iba a tu casa, pero mi mamá me pidió que fuera a recoger unas compras en el centro. ¿Quieren acompañarme?

Pablo: No oí bien. Alberto, ¿adónde va Luis?

Alberto: Dice que ___(1)___ al centro. Le ___(2)___ su mamá que ___(3)___ unas compras. Quiere saber si nosotros ___(4)___ acompañarle.

Student writes:

1. va
2. pidió
3. recogiera
4. queremos

Comment:

A modicum of ingenuity in the construction of the context makes it possible to avoid using multiple-choice items. The student's range of choice is severely restricted and yet he is required to perform more actively than he is when he recognizes and selects one of four alternatives.

(9) Language Area: Grammar units

Skill: Reading

Testing Device: Multiple choice

Instructions: For each of the following incomplete sentences check the word or phrase that best completes the sentence: meaning, grammar and style are all relevant considerations.

Student sees:

1. Se supone en España que el campesino _____ es más directo, sencillo, agudo y noble que el nacido en la ciudad y crecido en ella.

____ (A) sea
____ (B) está
____ (C) estará
____ (D) esté
__x__ (E) es

2. Estos son otros títulos traducidos _____ varios idiomas.

__x__ (A) a
____ (B) en
____ (C) con
____ (D) desde
____ (E) para

3. Las viejas _____ las palabras del joven.

__x__ (A) comentaron
____ (B) comentaron a
____ (C) comentaron de
____ (D) comentaron en
____ (E) comentaron sobre

Comment:

An answer sheet can be easily designed for items of this type. These items can be constructed only if enough good distractors can be found; it is generally not possible, therefore, to make very widespread use of them until the students have progressed well beyond simple Spanish.

(10) **Language Area:** Derivation

Skills: Writing/Reading

Testing Device: Completion

Instructions: In the second sentence of each of the following pairs there is an unfinished word. Fill in the blank with the missing part of the word. The meaning of the completed second sentence in each pair must be consistent with the meaning of the first sentence of the same pair.

Student sees:

1. Enrique nada muy bien.
   Enrique es muy bien nad____.
2. Vimos claramente cuánto nos lo agradecía.
   Su agradec_____ era evidente.
3. Su reacción fue violenta.
   Reac_____ violentamente.

Student writes (in the blanks):

1. ador
2. imiento
3. cionó

Comment:

This format requires the student actually to produce the appropriate elements. A multiple-choice item to test the same knowledge would have two drawbacks: (1) it would require the student merely to recognize the correct answer, and (2) it would necessarily suggest a number of non-existent forms (the distractors).

(11) Language Area: Syntax (Clauses/Sentences)

Skills: Writing/Reading (Listening comprehension/Speaking)

Testing Device: Transformation

Instructions: In each of the following items a question is followed by a series of words from which you are to construct an appropriate answer to the question. Your answer must be both grammatically and stylistically appropriate. You may add any words that are necessary to make complete sentences; do not add any unnecessary words.

Student sees:

1. ¿Dónde está tu hermana?

   estar / su cuarto / donde / vestirse / ir / con / yo / al cine

2. ¿Con quién hablabas cuando te vi?

   hablar / profesor / química / porque / salir mal / examen

3. ¿Qué quería tu papá?

   querer / yo / ir / comprarle / cigarrillos

Student writes (or says):

1. Está en su cuarto donde se viste para ir conmigo al cine.
2. Hablaba con mi profesor de química porque salí mal en el examen.
3. Quería que yo fuera a comprarle cigarrillos.

Comment:

By using a context of the kind that a question supplies, it is possible to control carefully even rather complex responses.

(12) Language Area: Vocabulary

Skill: Writing

Testing Device: Completion

Instructions: Complete the following list by filling in the blanks with appropriate words.

Student sees:

1. (male/female)

(a)	hombre	________
(b)	________	yegua
(c)	perro	________
(d)	toro	________
(e)	________	gata
(f)	rey	________

2. (antonyms)

(a)	bueno	________
(b)	________	bajo
(c)	________	despertarse
(d)	acostarse	________
(e)	delante	________
(f)	________	debajo

Student writes:

1. (a) mujer
   (b) caballo
   (c) perra
   (d) vaca
   (e) gato
   (f) reina

2. (a) malo
   (b) alto
   (c) dormirse
   (d) levantarse
   (e) detrás
   (f) encima

Comment:

This is one of the simplest and most direct ways to test vocabulary without resorting to multiple-choice items. (Pairs of sentences can be constructed to supply ample context in which the student must fill in a blank with the grammatically and stylistically appropriate word. However, it is time-consuming to write such items; it is therefore best to use them to test more advanced students and more subtle points than those illustrated in the examples given here.)

(13) Language Area: Style

Skills: Writing/Reading

Testing Device: Completion/Multiple-choice

Instructions: Complete the following dialogue with words or phrases chosen from the list to the right of the passage. Write your answers in the blanks provided. There are more words and phrases in the list than you will need. Some words or phrases may be used more than once.

(María y Elena son buenas amigas. Son mexicanas. Tienen 16 años de edad. Se encuentran por la calle. Luego se acerca a ellas un señor anciano de aspecto muy distinguido.)

María:	Buenos días, Elena. ¿Cómo _____?
Elena:	Muy bien, gracias. ¿Y _____?
María:	¿Adónde _____ con tantas prisas?
Elena:	Iba precisamente a _____ casa. Quería pedir _____ un favor.
María:	Bueno, _____ me lo que es.
El señor:	Buenos días, muchachas. Por favor, _____ me dónde se halla el Banco Nacional. Se me ha dicho que está por aquí.
María:	Está muy cerca, ______. Para llegar a la entrada principal del banco, _____ _____ derecho por esta calle. Está a unos cien metros de aquí, en este mismo lado.

decid
di
díga
dígan
está
estás
le
señor
siga
sigue
su
te
tu
tú
usted
va
vas

Student writes in:

estás - tú - vas - tu - te - di - dígan - señor - siga - usted

Comment:

The student must choose between tú and usted according to relevant social criteria, and he must choose between digan and decid on the basis of his knowledge of Mexican usage.

(14) Language Area: Literature

Skill: Reading

Testing Device: Matching

Instructions: Each of the authors in the left-hand column is generally associated with one of the choices in the right-hand column. Enter the letter corresponding to your choice from the right-hand column into the blank to the left of the number corresponding to each author.

___ 1. Benito Pérez Galdós

___ 2. José María de Pereda

___ 3. Emilia Pardo Bazán

___ 4. Juan Valera

(A) Naturalismo literario
(B) Elegancia de estilo
(C) Actitud conservadora general
(D) Actividad política de tendencia socialista
(E) Innovación en cuestiones religiosas
(F) La realidad vital de la vida española del siglo XIX

Answers:

1. F
2. C
3. A
4. B

Comment:

This item requires the student to associate a general characteristic with a writer; it demands more than mere recall of memorized titles, dates, etc. The authors are a homogeneous group of nineteenth-century novelists. Items of this type are difficult to write.

(15) Language Area: Culture

Skill: Reading

Testing Device: Matching

Instructions: Enter into the space to the left of each number the letter corresponding to the region with which each of the following products is associated; no region can be used more than once.

___ 1. pescado

___ 2. corcho

___ 3. vino

___ 4. naranjas

(A) Extremadura
(B) La Alcarria
(C) La Rioja
(D) Navarra
(E) Asturias
(F) Andalucía

Answers:

1. E
2. A
3. C
4. F

Comment:

This item tests the student's knowledge of Spanish agriculture, geography and associations that are frequent among Spaniards.

(16) Language Areas: Intonation/Grammar units

Skills: Speaking/Reading

Testing Device: Transformation

Instructions: You will hear a number of statements. In each case formulate a question to which the statement is an appropriate answer.

Student hears:

1. Va a Madrid.
2. Se fue con Paco.
3. Salió mal porque no había estudiado bastante.

Student says:

1. ¿Adónde va?
2. ¿Con quién se fue?
3. ¿Por qué salió mal?

Comment:

Written cues could also be used for an item of this kind. This item can easily be administered in a language laboratory. Items of this type take much time to grade; the teacher must spend hours listening to students' tapes. For that reason they must be written and administered with considerable forethought so that the information they yield justifies such an expenditure of time and effort.

CHAPTER V

The Interpretation and Use of Test Results

5.1 What is a Test?

Basically, the purpose of all testing techniques is to provide information for the teacher about his students' progress. There are many techniques the teacher can use to obtain information. Any technique is a 'test' to the extent that it yields the information the teacher seeks.

5.2 What is a Good Test?

A test is effective or ineffective only in relation to its particular function. For example, when a test is designed to evaluate mastery of subject matter it is not necessary that this same test also discriminate clearly between the slightly above average and the slightly below average students. A test designed to measure the student's ability to organize facts may not be an effective instrument for measuring his ability to recall information. A test designed for grading purposes may be good neither for motivating students nor for isolating weaknesses in teaching or student shortcomings. In other words, the purpose of the test should determine the content of the test, the form of the test questions, and the use that is made of the test results.

5.3 The "Quiz"

5.3.1 The quiz is typically brief in length and tests a relatively small body of material. The quiz is a somewhat more formal kind of continuing evaluation than day-to-day oral questioning of students in class.

Quizzes are used by teachers for two purposes. (1) They are used to encourage students to do their work regularly and to organize their knowledge systematically. (2) Teachers also use quizzes to judge the progress of a course; they confirm that the students have mastered specific facts or principles and reveal what is not clear and needs further elaboration.

5.3.2 Some teachers feel that quizzes should not be announced in advance or given at regular, predictable intervals because students will presumably study more regularly if a quiz is possible at any time. Others feel that quizzes should be given regularly and predictably so that students can prepare for them systematically. In general, unannounced quizzes tend to have punitive overtones and imply distrust of the students; the result is usually poor classroom morale.

5.3.3 In planning the classroom quiz the teacher must strive to insure that the students will recognize clearly the relationship between what they have studied and the content of the test. Analyses of classroom tests reveal that many teachers do not observe this principle.

Students are very quick to perceive a teacher's bias in testing. They naturally tend to study whatever they think will be useful for passing tests. If the teacher wants to encourage his students to master certain specific skills and concepts, he should be explicit about the nature and form of the questions to be used in the examination; he should not expect the students to infer the character of the examination from such clues as they are able to extract from the teacher's presentation in class.

5.4 The Unit Test

Most teachers subdivide the work of a term or semester into units or blocks of instruction organized around particular topics. The common theme in any particular unit should be evident to the student. It is important that the teacher make explicit the relative importance of different aspects of a particular topic so that his students will know how to budget their time effectively in preparing for an examination. His explanation may be nothing more than a simple list of topics or it may take the form of a more elaborate outline in which individual topics are broken down into subtopics. Unless he is told otherwise, a student is likely to assume that the amount of time spent on a topic in class directly reflects the importance of the topic for the teacher. The teacher should tell his students explicitly what the relative importance of each topic is for a test. A detailed plan of this sort can be drawn directly from the teacher's class outlines. The plan can also serve as a useful reminder for the teacher in preparing the test itself.

5.5 The End-of-Term Examination

The end-of-term examination is necessarily somewhat different from other tests. Evaluations made during the term (continuing evaluation in class, quizzes and unit tests) provide feedback both to the teacher and to the student. The end-of-term examination, however, is essentially an achievement examination that should reveal the extent to which the student has mastered the material of the course.

Some teachers do not explain clearly to the student exactly what it is they want him to learn in a course. It does not help him much for the teacher to say merely that one of the objectives of a language course is to develop in the student an appreciation of the culture of another country; the teacher must make explicit which particular aspects of that culture he will treat. Instructional objectives must be made sufficiently precise for the teacher to be able to decide whether the student understands what he has studied. In addition to making clear, both in his instruction and in his examinations, which particular aspects of the foreign culture, for example, he wants the student to master, the teacher must plan in advance a specific procedure for assessing the student's mastery of the material.

5.6 Length of the Test

5.6.1 Teachers frequently ask how long a test should be. Since the 50-60 minute class period has come to be a standard unit of time, the question may be restated to ask how many questions may be used in one class period. We might say that ideally the teacher should spend as much time evaluating each student as he needs to make a reliable judgment. This implies that different amounts of time might be spent testing different students. For example, for a student who has demonstrated his competence in the language itself on numerous occasions, the teacher needs relatively little time to verify the student's command of the language. Similarly, if a student has characteristically performed poorly in the language, only a relatively brief series of questions is needed to confirm that the student's performance has not improved. For the teacher, the most difficult problems in evaluating student achievement arise for students in the middle range of ability, or with students whose performance has been erratic. To make the most economical possible use of testing time it can be argued that the teacher should plan to spend only as much time with the very able and very inept students as he needs to confirm that they have followed a predicted pattern. He can then concentrate on those students whose behavior is contrary to his expectations and on students in the middle range of achievement.

5.6.2 However, most teachers do not individualize their testing; instead, they test all students in the same way. This procedure obviously has certain advantages of ease and simplicity of administration, and most teachers seem to be convinced that more individualized assessment is impractical. Whether or not more individualized assessment is impractical, in actual practice group tests for entire classes of students are the norm. For this reason, the remainder of this discussion will concentrate on group-testing procedures.

5.6.3 Various writers have suggested rules of thumb for determining the number of items of a given type that can be used in a specified number of minutes, but such guidelines are of limited usefulness. An example is the 'rule' that students can answer approximately one multiple-choice question a minute or three true-false questions per minute. In deciding on the number of items which students should be expected to consider in a given period of time, two important factors must be taken into account. The first of these is reading time. The reading time of a test is the amount of time required of the student to read and understand both the directions and the intent of each separate question. The second is the amount of time required for the student to respond to the question. If the student is familiar with the item type and has a thorough grasp of the subject being tested, the problem-solving process may occur almost simultaneously with the reading phase, especially when the items are relatively simple. However, when the student is expected to reformulate, reinterpret and synthesize prior experiences to solve new problems he will generally need additional time to respond to an item once he has read and understood it.

5.6.4 Teachers should also bear in mind that students vary in their approach to test questions. Some students, for example, pace themselves so that they consider at least briefly all the items a test contains, no matter how long it is, while others seem to need to answer each item satisfactorily before leaving it to go on to the next. There is no obvious correlation between a student's approach to a test and ability in a subject. We find both good and poor students moving rapidly through examinations and some of both kinds ponder and worry over thorny questions even at the price of not having time to consider all of the items in the test.

5.6.5 To some extent the students' approaches to tests can be modified by the teacher through appropriate phrasing of directions. Some students, however, seem unable to budget their time effectively in order to complete the test in the available time.

5.6.6 Research on the correlation between completing a test and rank in class seems to indicate that, within very broad limits, allowing enough time for all students to finish a test does not appreciably improve the standing of the slower students in the class. Experiments have been carried out in which examiners have stopped all students at the moment when the fastest student completed the test, scored the papers, and then allowed those students who wished to do so to spend more time to complete the examination; these experiments indicate that the relative standing (rank) of students in a class is not very different under these conditions. These findings have been replicated sufficiently often to indicate that the primary benefit of allowing a student as much time as he feels he needs to complete the examination is psychological, not statistical: the student who has had time to finish the test may feel better about it, even though he has not improved his standing in the class. Most teachers avoid examinations which are so long that the majority of students cannot complete them. In general, a test ought to be designed so that the fastest students will finish it in about three-fourths of the available testing time. The teacher's own experience with different types of items of his own making should guide him in deciding how many items and how many of each kind to put into a test.

5.7 Weighting the Examination

When the teacher prepares his own achievement tests, there are two ways in which he can reflect in them the relative importance of the various topics treated in the course. First, he may make the proportion of items treating a given topic in the test correspond to the importance of the topic in the course as a whole. Secondly, he may explicitly assign more points to some items than to others. The teacher may also include in the test some items which he judges to be more difficult than others and to which he assigns a larger number of points for that reason. He must be sure in any case that in the context of the whole test the total weight of all items related to a particular topic is proportional to the importance of the topic. However, the teacher should always bear in mind that he can estimate only more or less accurately the real difficulty of an item in a test; only the results of an actual administration of the test will present a clear picture of the relative difficulty of the items that make it up. Nevertheless, it is important in composing a test constantly to bear in mind the probable relative difficulty of the items in it.

5.8 Administration of Examinations

5.8.1 Many teachers are impatient with the details involved in reproducing, administering and mechanically processing test papers. However, much of the testing time can be wasted, both by the teacher and by the students, unless the test papers are clearly legible, the directions for the various parts of the test are clear and explicit, and everyone understands exactly how the test is to be administered and scored. For example, if the teacher has to interrupt the class while it is taking a test to point out typographical errors in the text of the test itself, the students' train of thought will be disrupted, and they may conclude that since the test was not prepared with great care they need not be careful in answering the questions in it. Students have a right to expect that tests will be carefully prepared and edited.

5.8.2 Directions for each part of a test should be clear and complete. The more familiar students are with a particular item type, the simpler the explicit directions for it can be. However, the realistic teacher is well aware that many students completely ignore printed or oral directions in their haste to begin to respond to the questions themselves. It is particularly important, therefore, that students be made aware of any unusual features of any part of a test before they begin to take the test itself. It is particularly important that young children who have had relatively little experience with written test materials understand exactly what they are to do. It is wise to use standardized directions for each separate item type; such standardized directions may be borrowed from the public domain or they may be evolved gradually by the individual teacher himself. The adequacy of a set of directions can be checked by submitting the directions and some sample items to students of modest ability; if directions are not clear to students of average or below-average ability when they are not being tested, they will surely not be clear to the majority of the students taking an actual examination. Unclear directions necessarily have an adverse effect on the individual student's performance; his performance on a test in which the directions are unclear is not a true index of his ability or his achievement.

5.8.3 A number of minor matters tend to be ignored in the administration of tests. Remember to bring extra paper if the student is expected to prepare a written response. Provide staples or paper clips in order to prevent the loss of papers. Remind the students to put their names on each sheet of paper so that if the pages become separated they may be properly collated. Decide beforehand on the arrangement of students in the room so that the test can be adequately proctored. Rehearse with the proctors what assistance they will be allowed to give students who are having difficulty. Arrange for adequate means of distributing and collecting the test papers so that the security of the test can be maintained. Each of these problems and many others can be awkward or irritating for the test administrator if he has not anticipated them ahead of time.

5.9 Forms in Which Test Results May Be Expressed

5.9.1 If testing is the collection of information upon which instructional decisions can be based, the kind of decision to be made should determine the form and the content of the test and the way in which the results of the test are expressed. The form in which the test score is reported should be appropriate to the purpose for which the test is being used.

5.9.2 For example, a teacher may conduct a question-and-answer session in class to find out whether the students know the past tense of the verb to be. A simple check-mark in a grade-book is sufficient to record the results of such an evaluation. Tests designed to elicit more complex responses from the student require more complex reports of the students' responses. In general, as the length of the test or of the individual responses which the student makes increases, more complex grades become necessary such as: superior, good, average, poor, or unacceptable; A, B, C, D, and F; or 1, 2, 3, 4, 5, etc.

5.9.3 In the case of objective tests (containing multiple-choice, completion, true/false items, etc.), each individual item is designed to measure something relatively specific. It is assumed, however, that such a test, taken as a whole, constitutes a complex evaluation of the student's ability or performance. The student's score is ordinarily the total number of questions answered correctly and is referred to as a raw score.[1]

5.9.4 By itself, a raw score (weighted or unweighted) for an individual student is relatively meaningless. To know, for example, that a student has a raw score of 32 points tells us almost nothing without further information either about the test (for example, the total number of points possible) or about the test performance of some appropriate reference group (such as the other members of the class). If we know, for example, that there are a total of 40 points possible on the test, we know that the student obtained 32/40 of the total number of points possible. If we set arbitrary standards about the proportion of the total a student should earn in order to get a grade of A or B or F, we can assign each student a grade based on his raw score. Under such a system of interpreting test results, it is possible that all students could receive an A.

---

[1] The assumption underlying this type of raw score is that all items are equally important and two students, each of whom has a raw score of 32, are judged equally competent, even though they may not have answered the same 32 items correctly. If, however, all items are not equally important, more than one point is assigned to some questions and a student's total score is obtained by adding together the number of points assigned to each question answered correctly. A total score based on this procedure is called a weighted raw score. On many tests weighted scores can be shown to provide a more satisfactory result than unweighted raw scores.

5.9.5 Where the teacher has designed the test with questions of different levels of difficulty (that is, with some items which only the A students should be able to answer correctly, another group of items which both the A and B students should be able to answer, etc.), we can say with some assurance that students who are able to answer from 35-40 items correctly should receive an A, students who answer from 30-34 items correctly should receive a B, and so on.

5.9.6 More often, however, tests are not designed with items of carefully gauged different levels of difficulty; instead, they are designed to cover certain material at a level corresponding to the average student in the class. Furthermore, many teachers are reluctant to set arbitrary standards (such as: 35-40 points is an A), if they are not sure that any student will achieve such a score. Teachers seem generally to prefer to relate each individual score to the scores of the other members of the class through the use either of rankings or of some kind of converted (standard) score. In effect, what most teachers want to know is not that a student answered 32 questions correctly, but how his score of 32 ranks in comparison to the scores earned by other students in the same class. If no one else in the class earns a score higher than 32, then 32 is regarded as an excellent score, even though the student has missed 1/5 of the questions in the test. Conversely, if most of the scores in the class are higher than 32, 32 is regarded as a poor score. For many teachers, then, the meaning of a raw score depends on the scores of other students in some reference group such as the remainder of the class. Once the teacher knows the rank of each student in the class, he can make a decision about the proportion of the class that should receive each letter grade.

5.9.7 In all cases, the distribution of grades, whether based on an absolute system (such as the proportion of correct answers) or a relative system (such as rank in class) involves certain arbitrary decisions by the teacher. No matter what grading scheme a teacher decides to use (numbers, letters, or descriptive phrases), he cannot avoid the responsibility for arbitrarily establishing "cut-off" scores or proportions of students in each grade or letter category. What he can and must do, by appropriate use of tests, is assign grades in a manner which is as unbiased and objective as possible.

5.9.8 The teacher will sometimes have several test scores or other grades which he wishes to combine into a composite evaluation of the student; in these cases the teacher should have a clear notion about the relative contribution that each separate piece of information should make to the composite description. For example, a teacher may have three unit tests and a final examination which he wishes to combine into a summary evaluation of a student's performance in an entire course. Generally, it is not possible simply to add up the total number of points a student has earned on the four examinations and divide that total by four to get an average, since four different tests will usually vary in length and importance. Let us suppose that the teacher decides that the three unit tests are equally important and that the final examination should receive the same weight as the three unit tests taken together. The simplest method is to give each student a rank on each test; the ranks for each student on the three unit tests are then combined and divided by three to obtain an average rank for the three unit tests. This average rank is then added to the student's rank on the final examination and this total is divided by two yielding a weighted rank[2] in which each of the unit tests is given equal weight, and the three unit tests together are weighted equally with the final examination. Such overall composite ranks can be arranged in order and grades assigned as if there were only one measure.

5.10 Test Statistics for Nonstatisticians

5.10.1 Descriptive statistics for use in test interpretation need not be complicated. As a case in point, consider the mean or average. In order to determine the exact mean of a set of scores the teacher adds the scores made by each student and then divides this total by the number of students in the class. The mean, however, is only one of the statistics used to represent a center point in a distribution of scores. While there are certain statistical advantages to using the mean, for purposes of test interpretation a better measure of central tendency is the middle score. The middle score, as its name implies, is the score earned by the student who ranks exactly in the middle of his class. In a group of 25 students, for example, the middle score is the score earned by the student who ranks 13th in his class. This value may not be exactly the same as the mean. The mean and middle score will usually not differ by more than a point or two. Without exception, the teachers will find the middle score entirely adequate for representing the center of a distribution of scores, and much easier to find than the mean.

---

[2] Technically, the values obtained are not ranks; however, the distinction does not alter the final outcome of the problem.

5.10.2 Let us now consider the standard deviation. Technically, the standard deviation is a measure of variation of the distribution of scores around the mean. The standard deviation is useful to the teacher for two reasons. First, the standard deviation of a test is a rough index of the reliability of the test. Other things being equal, a set of scores which is widely dispersed will provide more reliable descriptions of students than a test where all of the scores are tightly grouped around the mean or middle score. If scores are very close together small changes in performance, due perhaps to guessing or other chance factors, can substantially influence the order in which the scores appear. When scores are more widely dispersed, however, small variations due to such factors will be less influential in determining the arrangement of scores.

5.10.3 Secondly, the standard deviation is useful in interpreting test results when the teacher wants to compare the results of two or more tests. It can be shown that the weight (contribution) of a subtest in a composite score is proportionate to the standard deviation of that subtest. In other words, if we add together two tests one of which has a large standard deviation and the other a smaller standard deviation, the test having the larger standard deviation is more influential in determining the final order of scores than the test with the smaller standard deviation.

5.10.4 Fortunately there are useful approximations to the standard deviation which do not involve long calculations. One such approximation, which works surprisingly well, is found as follows: arrange the student scores in rank order, count down from the highest score to the rank which divides the upper one-sixth from the lower five-sixths and add up the scores in the upper one-sixth. Next, beginning from the lowest score, mark off the lowest one-sixth of the class and add up those scores. Then, subtract the sum of the scores in the lowest one-sixth from the sum of the scores in the upper one-sixth, multiply this difference by two (2) and divide the resulting product by the number of students. The resulting figure is a good approximation of the standard deviation. The estimate of the standard deviation obtained by this procedure is entirely satisfactory for most purposes of test interpretation. A second way to calculate an approximation of the standard deviation is to arrange the scores in rank order, count down one-sixth of the scores and record the score earned by the student one-sixth of the way down the distribution, then go to the bottom of the distribution, count up one-sixth and record the score earned by that student, and, finally, subtract the two scores and divide the difference by two.

5.10.5 A third statistic of some importance in test interpretation is known as the standard error of measurement. The standard error of measurement reflects the reliability of a test; it provides an estimate of the amount of variation in individual scores that would be obtained if the students were tested a second time. The standard error of measurement is useful in interpreting test results because it reminds the teacher that the particular score earned by the student depends both on the student's true ability and on unreliability of the test, guessing, and other factors not directly related to the student's ability or achievement.

5.10.6 The standard error of measurement provides an estimate of how much of the student's score is due to his true ability. It is an indication of the approximate range within which one would expect a student's score to vary from one occasion to another. For most teacher-made tests, the standard error of measurement is between five and six-tenths (.5 to .6) of the standard deviation of the test. For example, consider a test which has a standard deviation of six points. For this test the standard error would be from three to about three and a half points (0.5 x 6 = 3.0, 0.6 x 6 = 3.6). In other words, for most students, one would expect scores on a second administration of a test to be within 3 to 3.5 points above or below the score they made on the first. Although the teacher would still use the actual score earned by the student, the standard error of measurement should remind him that the score he has before him is based upon one sample of the student's performance and that with another sample the student's score might vary several points from the value actually obtained.

5.10.7 Statistics are an aid to interpretation. By calculating certain simple values from the distribution of scores, the teacher can obtain a great deal of information which is useful in describing both the performance of individual students and the performance of the class as a whole. The teacher who uses test results without employing any descriptive statistics relies on impressions which may or may not be justified, and he fails to use all of the useful information that is readily available to him from the test itself.

5.11 Test Reliability and the Teacher

5.11.1 In most texts on measurement, the concept of reliability is explained in such obscure and unfamiliar statistical terms that the teacher may not see what relevance it has to what he does in teaching and testing. It is unfortunate that reliability has been cast in this light since it is precisely in the context of classroom measurement that reliability is of most pressing concern.

5.11.2 For the teacher using published (standardized) test materials, reliability is a relatively simple matter. Any test which is marketed by a reputable publisher can be assumed to have adequate reliability if it is carefully administered and scored. In using standardized tests all the teacher needs to do is to read carefully the manual provided with the test and remember that test scores are always approximations rather than absolute values.

5.11.3 With teacher-made tests, however, reliability is an important question. It is important because the reliability of classroom measures can be greatly affected by the procedures used by the teacher and because evidence shows that teachers' judgments are often highly unreliable and thereby misleading.

5.11.4 Whenever we describe an object or event, what we describe does not stand in isolation but exists as part of a stream or continuum out of which we have taken certain elements: a sample. For example, we realize that a test intrudes on a student at some point in time, but we assume that the underlying behavior of the student is a continuous characteristic which is stable and predictable from one occasion to another. We do not expect that a student will be at the top of his class on one day and at the bottom on another, that a student will be able properly to conjugate one verb and not another similar verb, or that a student will be able to write an acceptable paragraph one day and not another. This assumption of continuity and stability is essential to any meaningful or useful concept of reliability. Without this assumption we would never be able to relate one event to another or to separate real changes in the student from unreliability in our descriptions or measurements.

5.11.5 If classroom measurements describe relatively stable characteristics of students, it is useful to be able to specify the extent to which a description (measurement) of student ability made on one occasion predicts a similar measurement or description made on another occasion. In other words, to what extent can we expect descriptions of a given individual to agree from one time to the next and from one observer (teacher) to another? Reliability is this property of a measure to be relatively constant through time, or the agreement by two or more observers on the same description. Measures and descriptions which do not have at least one of these characteristics are said to be unreliable.

4

5.11.6 Let us now explore some of the kinds of consistency and agreement which teachers might wish to find in their measurements or descriptions. A teacher would hope to be consistent in his description of a student's understanding of a given fact or principle. For example, we might like to be able to describe an individual student's ability to use the subjunctive and we might further like to place the student in one of three descriptive categories: (1) he can use and recognize the subjunctive with adequate facility; (2) he can usually recognize a subjunctive form and can sometimes use it appropriately; (3) he can rarely recognize or use the subjunctive correctly. In order to evaluate a given student's ability, we would select a number of items in which the student could demonstrate his ability to use the subjunctive. From this sample of instances we should like to be able to generalize about the student's ability to use the subjunctive properly. Because the teacher is dealing with only a sample of the student's behavior, it is important that this sample be representative of the larger class of behavior about which the teacher would like to generalize. In other words, we would like to be able to describe the student's ability to deal with the subjunctive in such a way that our description would not be invalidated by another sample of items. We are concerned with the stability of our description (and the student's performance) from one set of items on the subjunctive to another.

5.11.7 A second kind of stability which should be of concern to the teacher involves the consistency of his own judgments on two or more occasions. Let us consider the case of a teacher reading a set of student papers. By chance, he happens to read two very good papers and then reads a poor paper. To what extent might the teacher's judgment about the third paper be influenced by what he has read in the first and second papers and how different might his judgment of the third paper have been had it been preceded by two equally poor papers? Ideally, we might hope that the order in which the papers are read will not affect the teacher's judgment. Research indicates, however, that judgments are made in a context and that events immediately preceding a judgment do influence a judge's decision. We also find that teachers have considerable difficulty judging objectively the behavior of students for whom they have made prior judgments on other occasions. As teachers, we have all been tempted to let our prior experience with students override new impressions that do not agree with our preconceived notions. Who has not given a capable student the benefit of the doubt in cases where the student's response to a question is not complete or entirely clear? Conversely, which of us has not resisted the tendency to change his opinion of a student on the basis of one or two favorable impressions scattered among a large number of unfavorable ones?

5.11.8 It is obvious that we could greatly increase the consistency of our judgments or descriptions of an individual student by taking into account our previous experience of him. But consistency due to prior knowledge constitutes bias. Our real concern is to understand what consistency would result if our judgments were made completely independently of one another.

5.11.9 A third type of consistency which is of concern to the teacher is the degree to which two or more observers (teachers) would agree in their description of a given student's behavior. It seems reasonable to suppose that two competent, independent observers should agree on their description of a student's behavior. The evidence suggests, however, that unless independent observers have been trained to look for similar things in the student's behavior and unless they have been given a common set of descriptive categories, agreement among them will not be high.

5.11.10 In each of these types of consensus or agreement, reliability is important to the teacher: (1) consistency of a particular student's behavior from one sample to another; (2) consistency of a given teacher's judgment on two or more occasions; (3) agreement among teachers concerning the description of a particular student's ability or performance.

5.11.11 Let us consider ways of improving the reliability of such descriptions. A teacher wants to know, for example, whether a student can distinguish between two vowel sounds. He selects a large number of words in which the vowel sounds occur and draws a sample of such words to be read to the student. He needs to know how large a sample of words to use to distinguish between students who can and those who cannot hear the distinction. If he uses only one word, there is a very good chance that a student who is not generally able to make the distinction will simply guess the correct answer. Conversely, there is also the possibility, for a variety of reasons, that a student who is generally able to hear the distinction may miss the sound on a single, specific word. If we include a second item, we will not expect the inept student to guess correctly on both items and, similarly, we will not expect the able student to miss the vowel sound in both words. Therefore, pairs of items are more reliable than single items, and the addition of further items adds still more reliability to the description. At some point, however, additional items result in very little increase in the reliability of the description.

5.11.12 The two primary factors which affect the reliability of a test item are the difficulty of the item and the relationship of the item to other items in the test. It is obvious that if a test item is so easy that every student answers it correctly, it cannot discriminate between good and poor students. Similarly, if an item is so difficult that no student answers it correctly, it is also useless for this purpose. One may include several easy items in a test to motivate the student, but such items have no value in terms of reliability. Items of moderate difficulty prove to be the most discriminating (and therefore the most reliable in this sense). The optimum value for item difficulty (the proportion of a group answering the item correctly) falls in the area where between one-half and three-fourths of the class answers the item correctly. The second characteristic of an item which influences the reliability of the test is its relationship to other items in the test. In general, the more narrowly focused the content is of a group of items in a test, the more reliable the test will be. Conversely, the broader the content included in a test, the greater the number of items that are required for useful reliability.

5.11.13 Reliability of teacher judgments on two or more occasions can be estimated in several ways; in its simplest terms it may be regarded as a problem in ranking. Let us consider a class of 20 students who have been arranged in rank order from best in the class to poorest (ranked 1, 2, . . . , 20). We would like to know the degree to which these ranks would remain the same if, using the same information, the same teacher ranked the students again (without, of course, referring to his previous rankings). Experiments show that the agreement among rankings is quite high at the extremes. In other words, with relative consistency, teachers can pick out the most able and the least able students in a group. Near the middle of the group, however, the differences between students become less obvious and the rankings on successive occasions become more variable. When a teacher bases his rankings on objective records, his reliability increases. On the contrary, when his rankings depend on more subjective judgments, reliability decreases. This is not to say that subjective judgments must be ignored in ranking students; however, when such data are used, judgments must be made as objective as the material will permit.

5.11.14 In the reliability of rankings, what degree of precision can reasonably be expected? To pursue our example further: it is unreasonable to expect a teacher to be able to reproduce two sets of 20 rankings with complete rank-by-rank agreement. However, it is reasonable to expect a very high degree of agreement if the teacher is asked only to group the students into upper quarter, lower quarter, and middle half. The fact that a given student might be ranked first on one occasion and third on the second occasion would not move the student out of the upper quarter. Similarly, a shift in rank even from 6th to 15th would still leave the student in the middle half of the group. This second principle has clear implications for the reliability of the teacher's judgments. When ranking students, it is rarely important to be able to put students in 15 or 20 different categories, since most decisions (such as grades) are based on only four or five categories. The majority of any class of students can be placed in three, four, or five categories with little difficulty, although there will always be a few students whose performance cannot be so neatly categorized. For those few cases, the teacher may either turn to measures which are entirely objective or to additional factors which are entirely subjective, such as attitude or motivation (with the justification that, when all achievement factors are equal, the interested or cooperative student should be encouraged).

5.11.15 The third type of consistency which is of concern to teachers is agreement among professionally competent observers. Whenever teachers' judgments are intended for use outside the classroom (as is the case, for instance, for term grades), it is important that such judgments represent the behavior of the student in some readily understandable way. If the teacher records student achievement for his personal use, it makes little difference what system he uses. Term grades, however, are a medium of communication whereby the teacher informs the student, his parents, other teachers, administrators, college admissions officers, and potential employers of the degree to which a student has met the objectives of a certain unit of instruction. Research on agreement among teachers indicates that when teachers specify clearly to one another what it is they wish to evaluate and when they define their rating scale carefully, they can achieve very high levels of agreement. Without such a common understanding, it is only by coincidence that two teachers will assign similar grades to a class of students.

5.12 Validity of Classroom Tests

5.12.1 By definition, a classroom test is valid if it provides the teacher with information which he finds useful in making instructional decisions. It is valid in the same way that a ruler is a valid measure of length or a scale is a valid measure of weight.

5.12.2 The problem concerning the validity of classroom tests resides in a possible lack of agreement between what a teacher actually measures and the kinds of abilities and skills he demands of the student. Just as we would question the validity of calling measurements made with a ruler "weights," we can question the validity of measuring "reasoning" by a test made up predominantly of highly factual items. The question of validity of classroom tests is posed when the teacher thinks he is measuring one ability and other, objective observers (including his students) believe he is measuring something else.

5.12.3 To illustrate the problem, let us consider a teacher who professes to use an audio-lingual approach to language instruction in his classes and then employs only paper-and-pencil tests to examine his students. Although it can be argued that certain aspects of audio-lingual ability can be tested in written form, to carry the argument to an extreme and use no other type of item would raise certain questions about the validity of tests purporting to measure audio-lingual skills in general.

5.12.4 A classroom test is invalid, then, only to the extent that a teacher thinks it is measuring one specific ability while it is, in fact, measuring something quite different. It is not difficult to diagnose such discrepancies. (Students are very quick to notice such discrepancies and tend to have little hesitation in deciding how to expend their time and effort. Regardless of what the teacher may say about his emphasis in a course, his students will be guided by the kinds of tests he uses.) All a teacher needs to do in order to know what his test is measuring is to sit down with a few students and ask them (1) how they studied for the test and (2) what they think specific questions mean.

5.13 Item Analysis

5.13.1 Many of the discussions of item-analysis techniques which appear in the literature on testing were not written primarily for the classroom teacher. These discussions often do not recognize that most teachers have not been shown how item analyses can contribute to their effectiveness as teachers. Nevertheless, if the teacher can readily obtain information of this kind about items on his examinations he will find it extremely useful.

5.13.2 An item analysis can be broadly divided into two phases. Phase I consists of identifying the statistical characteristics of items. The second phase involves interpretation of these statistics to determine which items serve their intended purposes and which items need revision.

5.13.3 Let us look at two very simple procedures for collecting the data necessary for an item analysis. The first of these procedures involves using the students themselves to help summarize the individual item responses. After the teacher has scored the papers, he arranges the papers into three groups. Group 1 should contain the papers with the highest one quarter of the test scores Group 2 the middle 50 percent, and Group 3 the lowest quarter of the class. (The teacher should put a piece of masking tape over the name of the student or otherwise conceal the identity of each student during the item analysis.) The papers are then marked "1", "2", and "3", to indicate which group they are in, and distributed to the students. Each student now has a paper (probably not his own) to be used in the following tabulation procedure. The teacher asks how many papers marked "1" have the first item answered correctly. The students raise their hands if the answer on the paper they are holding is correct. The teacher counts the number of hands and makes a tally, and proceeds to do the same for the papers marked "2" and "3". After he has gone through the entire examination in this way, the teacher has a tally of the number in the upper 25 percent of the class who answered each item correctly and the number in the lower 25 percent who answered the item correctly. The teacher may also have discussed the items with the class while making the tabulation, thereby using the time both to collect the item statistics and to review the examination with the students.

5.13.4 If an item is to discriminate among students according to their ability, we should expect that the more able students--those receiving the highest total scores on the test--will answer the item correctly more often than the less able students. In other words, a larger portion of the students in the upper 25 percent of the class should answer the item correctly than in the lower 25 percent. If the number of students who answer the item correctly in the two groups is approximately the same, then the item is said to be nondiscriminating: it does not discriminate between students according to their general ability (as measured by the total score on the test). If, on the other hand, the more able students do tend to answer the item correctly more often than the less able students, the item is said to discriminate positively. In general, positive discrimination is a desirable characteristic of an item; it tends to increase the range of scores and the reliability of scores. When the difference between the number of students in the upper 25 percent and those in the lower 25 percent of the class is negligible (or favors students in the lower 25 percent), an item is described as nondiscriminating (or negatively discriminating) and should be very carefully reviewed in the second phase of the item analysis.

5.13.5 From the number of correct responses in the upper and lower quarters of a class we can also obtain an index called item difficulty. This index is obtained by converting the number of students in groups 1 and 3 together into a percentage. This number, the item difficulty index, is a statement of the proportion of the whole group of students who answered the item correctly. In general, item difficulty indices should fall in the middle range (i.e., between 50 and 60 percent). A small number of relatively easy items near the beginning of a test can be defended because they help build a student's confidence, and a number of rather difficult items near the end of the test can be defended because such items challenge the most able students. More will be said about item difficulty in the treatment of the second phase of item analysis.

5.13.6 A second and less elaborate procedure for obtaining approximate estimates of the item difficulty and discrimination indices of a classroom test can be carried out entirely by the teacher, particularly where classes are small (between 25 and 30 students). The teacher goes through the papers of the most able students in the class and tallies whether or not they gave the correct answer to each question. He also goes through the papers of the poorest students. In phase I of the analysis the teacher identifies items which need further scrutiny. In phase II he examines them carefully.

5.13.7 If each item in an examination is included for some particular purpose, an item analysis should show whether it served the purpose for which it was included. If an item was included near the beginning of a test in order to build the student's confidence, the item was presumably intended to be easy. Such items, therefore, should be answered correctly by most or all of the students in the class and missed only by obviously incompetent students. If the item statistics indicate that this is how an item performed, the item should be judged to have served its purpose and therefore to be a good item. Similarly, if the teacher has included some very difficult items near the end of the test, he should expect these items to be answered correctly by only a few very able students. If the item statistics indicate that this was the case, then these are also good items.

5.13.8 The more difficult items to evaluate are those which fall into neither of these obvious categories and which make up the bulk of the test. In general, teachers should judge items according to two criteria. First, some items are unexpectedly easy or difficult for the class. An item which is identified as easier than the teacher had anticipated may indicate that the teacher has misjudged his students' abilities. This innacurate judgment may be the result of a lack of perceptiveness on the part of the teacher, it may be the result of the students' having prepared themselves particularly well for the examination, or it may be the result of some technical flaw in the item which allowed the student to choose the correct response without really understanding the intent of the item. The teacher should attempt to discover what the reason is for the unexpectedly good performance of the students. If he decides that the explanation is that he misjudged the students or that they prepared unusually well for the examination, all he need do is record his conclusion. However, he should be especially sensitive to the possibility that the students were able to answer the item because of scoring procedures which were not sufficiently rigorous or some technical flaw in the item itself. If technical or scoring problems appear to be the explanation for unusually high scores on an item, the teacher may wish to rescore the item (if it is a constructed response) or modify certain response options (if the item is of the selected response form) for future use. If the item appears to have been unusually difficult, a parallel set of explanations is possible. The teacher may have misjudged the students' ability. The students may not have anticipated that the teacher would ask certain questions and may therefore have ignored certain topics in their preparation for the examination. (Either of these explanations suggests insufficient communication between teacher and pupils.) There may be some mechanical problem in the way that the item is presented or scored.

5.13.9 Let us now turn to discrimination. The effect of nondiscriminating items is to make the range of scores smaller and thereby diminish the teacher's ability to distinguish between one student and another in terms of their ability or achievement. Items which do not discriminate between the most able and the least able students should be examined very carefully in an attempt to decide why they do not discriminate. In multiple-choice items, lack of discrimination is most often the result of poorly conceived distractors or poorly phrased correct responses. In the ideal multiple-choice item the correct answer seems obviously correct to the more able students while one or more of the distractors appear to be correct to the less well prepared students. To prepare the distractors, the teacher should draw on his experience of the kinds of errors students typically make with the point being tested. In constructed-response items the most common explanation for lack of discrimination is the tendency of teachers not to use the full range of scoring options available to them. If the teacher is using a 1-2-3-4-5 scale for scoring short-answer questions and he assigns all 3's and 4's to the students, such items will obviously not be very discriminating. Before he begins the actual scoring of such items, the teacher will find it useful to scan the papers to form a general idea of the range of the achievement represented in the responses. He should then decide on some allocation of points so that he will use the entire range of scores. If the student responses are so homogeneous that the teacher cannot make useful discriminations, the item probably needs substantial revision or should be eliminated altogether.

5.13.10 Occasionally a teacher will observe that the pattern of right and wrong answers to a given question seems to bear no relationship to overall test scores. He may find, for example, that many of the most able students unexpectedly missed the item and many of the least able students unexpectedly answered the item correctly. If such a pattern of responses to an individual item emerges, it should serve as a warning to the teacher that there is something unusual about this item and that it bears further special attention. It is possible that there is nothing wrong with the item other than that it happens to measure some ability or understanding which is different from those measured by the majority of other items in the test. More probably, however, it is simply a poor item.

5.13.11 Item analysis has two phases. Phase I involves the collection of item statistics to identify items in need of further attention. Phase II involves examination of these items. The implicit assumption underlying any analysis of test items is that the teacher intends to improve the quality of his testing by utilizing information obtained from each administration of his items. This in turn suggests that teachers should build their own files of items which have worked well, and that they should attempt to determine how they can improve their test-writing style. In the individual item responses the teacher may also find information about particular students' abilities or deficiencies which may be useful in planning further instruction.

5.13.12 Many teachers consider only overall test scores. There is, however, a great deal more information implicit in the results of the administration of a test than simply the single overall score. Let us say, for example, that among the items in a test the teacher has deliberately included several items which he feels every student should be able to answer. After the examination has been administered the teacher goes back through the students' papers to see whether they have actually mastered these points. He may examine the responses of an individual student who has not done well on the test as a whole to see which of the items dealing with basic information or basic concepts the student answered incorrectly. This type of detailed analysis is particularly useful when among the items included in a test there are some that test basic knowledge that is essential for a student to move to the next level of instruction. Examination of such minimum performance items allows the teacher to plan his teaching better than he can if he uses only the single overall test score. If the items in the test have been carefully planned, it is also possible to diagnose subtle difficulties of better students.

5.13.13 A substantial part of what is called item analysis is based on the intuition and judgment of the teacher. The statistical procedures which underlie item analyses can help the teacher to identify items which have unusual characteristics, but the final explanation of why an item works or does not work is largely a matter of the judgment of the individual teacher. With classroom tests very elaborate statistical analyses are generally not warranted. The teacher can usually identify those items which are working well and those items which are working poorly by judiciously selecting the better papers and the poorer papers and examining the response patterns represented in these two extreme groups.

GLOSSARY

The following are brief explanations of some technical terms as they are used in this Handbook.

Chance score: the score that can be obtained by guessing alone.

Contrastive analysis: the procedure whereby two languages are compared to specify the ways in which they are different.

Consonant cluster: a succession of consonants which is difficult to pronounce for non-natives (e.g., Italian /sb/ in sbaglio).

Control: formulation of a question so as to limit the ways in which it can be answered.

Corrected true/false item: a true/false item in which the key part of the stimulus statement is underlined. The student is directed specifically to use that crucial element to decide whether the statement is true or false.

Converted score: standard score; a general term referring to any of a variety of "transformed" scores in terms of which raw scores may be expressed for reasons of convenience, comparability, ease of interpretation, etc.

Diagnosis: use of test results to isolate those points which need further elaboration either for a few individual students for for the class as a whole.

Discriminating power: the ability of a test item to differentiate between good and poor students.

Distractor: an incorrect answer furnished in a multiple-choice question; the ideal distractor shares all of the features of the correct answer but the one being tested.

Distribution of scores: a tabulation of scores from highest to lowest showing the number of individuals that obtain each score or fall below it.

Global testing: the use of questions which do not attempt to focus on one specific learning problem and that, consequently, are likely to involve more than one skill, language area and testing device at the same time.

Halo effect: a bias in ratings arising from the tendency of the rater to be influenced in his rating of specific traits by his general impression of the person being rated.

Inflection: changes in words according to number, person, case, etc.

Item analysis: examination of test results to obtain data on score distribution, test reliability, etc.

Language areas: various aspects of a foreign-language curriculum (e.g., phonology, morphology, syntax, vocabulary, style, literature, culture).

Language feature: a specific element of a language where errors can be detected and corrected.

Language skills: the ways in which language competence can be displayed (listening comprehension, speaking, reading, writing).

Linguistic model: a method of explaining or describing a language according to a specific theoretical orientation (e.g., the structural model, the transformational model, etc.).

Modified true/false items: true-false items designed to reduce guessing by offering a choice among *true*, *uncertain*, *false*, etc. (see also *corrected true/false items*).

Morpheme: basic unit of grammar; the smallest meaningful unit of language. (It may be a word, or part of a word; *unfriendly* has three morphemes: *un*-, *friend*, -*ly*).

Morphology: the study of the classification of morphemes and the ways in which they combine to form words.

Oral-response test: a test in which the student speaks the answers directly to the teacher or into some sort of recording device.

Percentile: a point (score) in a distribution below which falls the proportion of cases indicated by the given percentile (the 15th percentile denotes the score below which 15 percent of the scores fall).

Phoneme: a distinctive class of sounds; the smallest unit of speech that distinguishes one utterance from another.

Phonology: the branch of linguistics that deals with the sounds of a language and their permissible combinations.

Positively discriminating: said of a test item if a higher proportion of good students than of poor students answers it correctly.

Raw score: the total number of questions answered correctly. (The assumption underlying this type of score is that all items are equally important; two students with the same raw score are judged equally competent, even though they may not have answered the same items correctly.)

Reliability: the consistency with which a test measures, or the degree to which repeated measurement of the same individual would tend to produce the same result.

Response set: a pre-disposition to answer questions in a certain way if there is uncertainty as to the correct answers.

Specific determiners: words giving clues which enable the student to answer an item correctly without possessing the specific knowledge desired.

Standard score: see *converted score*.

Stem: in a multiple-choice item, an incomplete sentence which is followed by several completions of that sentence among which the student is to choose the suitable one.

Stimulus: a brief statement, question or phrase to which the student is asked to respond either orally or in writing.

Syntax: arrangements of words in larger constructions, e.g., phrases, clauses, sentences.

Testing device: any of the various techniques used by the examiner to elicit the student's response (e.g., multiple-choice, true/false, or completion items, essay, dictation, etc.).

Validity: the extent to which a test measures what its author intends it to measure.

Weighted raw score: the total score which results when each item is assigned a number of points commensurate with its relative importance as compared with other items on the same test.

* * * * *

Bibliography

Books, booklets, and articles are numbered consecutively, although they are grouped according to the following categories:

I. General (Publications that offer an overview of testing in terms of the whole educational context.)

II. Testing and the Foreign Language Profession (Publications that deal with testing in terms of the professional context.)

III. Measurement Theory and Applications

IV. Foreign Language Testing

V. Statistics and Interpretation of Evaluation Results

    A. General

    B. Norms, Scores, and Their Interpretation

VI. Bibliographical Sources

Within each category or sub-category, books, booklets, and articles are sub-classified as "Introductory" and "Advanced or Technical." Those included in the latter category are marked with an asterisk.

I. General

1. Chauncey, Henry and John E. Dobin. Testing: Its Place in Education Today. New York: Harper & Row, 1963. xi, 223 pp., index.

   Appendix: "Multiple-Choice Questions: A Close Look." A reprint of parts of a booklet (see item #17) published by the Test Development Division of Educational Testing Service. Gives information on what is done with the multiple-choice form of question in many present-day tests.

   "This book is intended to provide teachers, interested school board members, and parents with a broad picture of what the testing in schools and colleges is all about . . . It is an overview of current educational testing developments, intended to contribute to a general education about education . . . The teacher who cannot become a specialist must . . . gain enough understanding of a particular field to be able to place it properly in the total spectrum of educational effort and to use its resources constructively and sensibly when the needs arise." (From the Introduction.)

2. Hoffmann, Banesh. The Tyranny of Testing. New York: Crowell-Collier, 1962. 223 pp.

   "An indictment of testing addressed to the general public. Contends that objective tests reward superficiality, ignore creativity, and penalize the person with a probing, subtle mind. Its nineteen chapters examine commonly used aptitude and achievement tests, and provide examples of test items the author considers faulty " (p. 24 of item #98 in this bibliography).

II. Testing and the Foreign Language Profession

3. Bryan, Miriam M. "Implications for Higher Education: Results on the New FL Tests." The DFL Bulletin (Publication of the Department of Foreign Languages of the NEA), V (May 1966), 1-3.

4. ---- "MLA Foreign Language Proficiency Tests for Teachers and Advanced Students." The DFL Bulletin (Publication of the Department of Foreign Languages of the NEA), V (Oct. 1965), 4-7.

5. "Guidelines for Teacher Education Programs in Modern Foreign Languages: An Exposition." Compiled by F. André Paquette, The Modern Language Association of America, The Modern Language Journal, L (Oct. 1966), 323-425.

   Reprint available from the Materials Center, Modern Language Association, 62 Fifth Avenue, New York 10011. Price: $1.00

6. MLA Bulletin of Information: 1967-68 MLA Foreign Language Proficiency Tests for Teachers and Advanced Students. See Item #61.

   Available free of charge from: MLA Foreign Language Proficiency Tests, Educational Testing Service, Princeton, New Jersey 08540

7. "Qualifications for Secondary School Teachers of Modern Foreign Languages." *Publications of the Modern Language Association of America*, LXX, 4, Part 2 (Sept. 1955), 46-49. Revised: *PMLA*, LXXVII, 4, Part 2 (Sept. 1962), 31-37.

   The "Qualifications" have been reprinted many times. They appear in items #5, 6, 9, 38.

8. Starr, Wilmarth H. "Proficiency Tests in Modern Foreign Languages." *Publications of the Modern Language Association of America*, LXXVI, 2 (May 1961), 7-11.

9. ---- "MLA Foreign Language Proficiency Tests for Teachers and Advanced Students." *Publications of the Modern Language Association of America*, LXXVII, 4, Part 2 (Sept. 1962), 31-42.

III. *Measurement Theory and Applications*

10. *College Testing: A Guide to Practices and Programs*. Prepared by the Committee on Measurement and Evaluation of the American Council on Education. Washington, D.C.: American Council on Education, 1959. x, 189 pp., bibliography.

11. *A Description of the College Board Achievement Tests*. New York: College Entrance Examination Board, Revised annually. Approx. 130 pp.

    Available free of charge from: College Entrance Examination Board, Box 592, Princeton, New Jersey 08540.

12. *A Description of the College Board Scholastic Aptitude Test*. New York: College Entrance Examination Board, Revised annually. Approx. 50 pp.

    Available free of charge: See Item #11.

13. Bloom, B. S., ed. *Taxonomy of Educational Objectives, Handbook I: Cognitive Domain*. London: Longmans, Green, 1956.

    An extremely useful handbook for helping teachers classify item types according to the cognitive process required of students to answer the item. The book also includes numerous examples of item types illustrating particular educational objectives. (Lathrop.)

14. *ETS Builds a Test*. Princeton, N.J.: Educational Testing Service, 1965. 24 pp.

    Available free of charge from: Evaluation and Advisory Service, Educational Testing Service, Princeton, New Jersey 08540.

15. English Testing Guidebook, Part II: Fundamentals of Test Construction and Interpretation. By David P. Harris. Prepared at the American University Language Center for the International Cooperation Administration. 43 pp., Selected References.

    Appendix: Descriptions of A(merican) U(niversity) L(anguage) C(enter) Tests.

    A very useful booklet.

    Available from: American Language Institute, Georgetown University, Washington, D.C. 20007.

16. Making the Classroom Test: A Guide for Teachers. Evaluation and Advisory Service Series No. 4. 2nd ed. Princeton, N.J.: Educational Testing Service, 1961. 27pp.

    Offers practical suggestions and illustrations for the improvement of teacher-made tests.

    Available free of charge: See item #14.

17. Multiple-Choice Questions: A Close Look. Princeton, N.J.: Educational Testing Service, 1963. 43 pp.

    Parts of this booklet are reprinted as an Appendix to item #1.

    Available free of charge: See item #14.

18. Katz, Martin. Selecting an Achievement Test: Principles and Procedures. Evaluation and Advisory Service Series No. 3. 2nd ed. Princeton, N.J.: Educational Testing Service, 1961. 34 pp.

    Helps the test-user to analyze standardized achievement tests and match them to his testing needs and purposes.

    Available free of charge: See item #14.

19. Stecklein, J.E. and R.L. Lathrop. Bulletins on Classroom Testing.

    A collection of 12 bulletins on various topics of practical interest to classroom teachers: What is a good test; How to prepare and evaluate various types of items; Grading and marking; etc.

    Individual bulletins or the entire set may be obtained from The Bureau of Instructional Research, University of Minnesota, Minneapolis, Minn. (Lathrop.)

20. Tyler, Leona E. Tests and Measurements. Foundations of Modern Psychology Series. Englewood Cliffs, N.J.: Prentice-Hall, 1963. xi, 116 pp., index, Selected Readings.

"This book has grown from my attempts to teach the basic principles of psychological measurement to people who do not consider these principles to be of central importance in their life and work . . . In my attempt to meet this need, I have found it useful to distinguish between producer and consumer knowledge. I have deliberately omitted or de-emphasized many concepts and methods that are essential in carrying out research or constructing standardized tests. And I have deliberately stressed the concepts one needs in order to read a journal article, choose a test to serve a particular purpose, or interpret an individual's score. . . My primary aim has been to make the basic ideas clear." (From the Preface.)

* 21. Adams, Georgia Sachs. Measurement and Evaluation in Education, Psychology, and Guidance. In consultation with Theodore L. Torgerson. New York: Holt, Rinehart and Winston, 1965. xiii, 654 pp., bibliographies, index.

Appendices: A. A Selected List of Standardized Tests for the Elementary and Secondary Schools; B. Publishers of Standardized Tests; C. Methods of Expressing Test Scores (Based on the Normal Curve); D. Selected Tables; E. Glossary of Symbols and Terms Used in Measurement and Evaluation.

"In this volume, almost every chapter focuses on the development of concepts. By attaining a thorough understanding of these concepts, students will be helped to select or develop measurement procedures appropriate to their purposes and to interpret measurement data with due respect for (1) the inevitable errors involved in any sampling procedure and (2) the limitations of indirect approaches typically used as efficient substitutes for direct study of behavior." (From the Preface.)

* 22. Anastasi, Anne. Psychological Testing. Second Edition. New York: Macmillan, 1961. xiii, 657 pp., author index, subject index, bibliographies.

Appendix: Test Publishers.

"An excellent introduction to the principles of psychological testing and to the major types of tests in current use. Approximately the first quarter of the text is devoted to a description of the uses and characteristics of psychological tests in general and to fundamental test statistics. The remainder reviews critically the classification, aptitude, achievement, and personality tests now available." (David P. Harris.)

* 23. Furst, Edward J. Constructing Evaluation Instruments. London: Longmans, Green, 1958. xv, 334 pp., index. Each chapter contains a list of References and another of Further Readings.

"In writing this book, I have intended it mainly for teachers and prospective teachers in our schools and colleges. . . . The major purposes of this book are to help such persons (a) acquire an understanding of the several basic problems involved in developing any evaluation technique, (b) further their understanding of principles of test construction, (c) become familiar with some of the best references in the field, and (d) broaden their understanding of the purposes which can be served by tests and other techniques of appraisal." (From the Preface.)

* 24. Gerberich, Joseph Raymond. Specimen Objective Test Items: A Guide to Achievement Test Construction. London: Longmans, Green, 1956. ix, 436 pp., indices, glossary, bibliographies.

"This book is designed for use by both teachers and test specialists. Its primary purpose is to serve the classroom teacher as a guide to achievement test construction. Its secondary purpose is to provide for the test specialist several systematic classifications of instruments and techniques used in achievement testing." (From the Preface.) Covers all subjects, very detailed bibliographies.

* 25. Ghiselli, E.E. Theory of Psychological Measurement. New York: McGraw-Hill, 1964.

An intermediate to advanced treatment of test theory for the reader who is a serious student of measurement and has a firm grasp of basic statistics and mathematics through algebra. The discussion is particularly helpful in relating individual item statistics to over-all test characteristics. (Lathrop.)

* 26. Helmstadter, G.C. Principles of Psychological Measurement. New York: Appleton-Century-Crofts, 1964.

A second level discussion of basic concepts used in measurement including reliability, validity and standardization. The treatment presumes no advanced mathematical sophistication and, although brief in some sections, is very readable to the person interested in going slightly beyond the very elementary discussions of measurement principles. (Lathrop.)

* 27. Lindquist, Everet F., ed. Educational Measurement. Washington, D.C.: American Council on Education, 1951.

"The comprehensive reference work and textbook on the theory and techniques of educational measurement. Twenty authorities on testing have contributed chapters in the areas of their specialization. The level of difficulty varies from chapter to chapter, and many sections are most definitely not for the beginning student of tests and measurement." (David P.Harris.)

An important collection of statements of particular interest to the serious student of testing. Although now somewhat dated many of the chapters on planning and building tests are still very useful to the classroom teacher. The chapters vary in difficulty and style. The volume, however,still represents one of the best collections of statements on measurement available in a single volume. (Lathrop.)

* 28. Remmers, Hermann Henry. A Practical Introduction to Measurement and Evaluation. New York: Harper & Row, 1960.

* 29. Thorndike, Robert L. and Elizabeth Hagen. Measurement and Evaluation in Psychology and Education. Second Edition. New York: John Wiley & Sons, 1961. viii, 602 pp., index. Each chapter contains bibliographical references and questions for discussion.

Appendices: I. Computation of Square Root; II. Calculating the Correlation Coefficient; III: Section A--General Intelligence Tests. Section B--Aptitude Test Batteries. Section C--Reading Tests. Section D--Elementary School Achievement Batteries. Section E--High School Achievement Batteries. Section F--Interest Inventories. Section G--Adjustment and Temperament Inventories; IV. Sources for Educational and Psychological Tests.

An introductory but rather technical text. "We continue in the basic belief that it is as important for students to learn what tests will not do as to learn what they will do; as important to examine their own purposes and objectives for testing as to examine the tests. It is in the hope of developing more restrained, discriminating, and insightful testers that we offer this book to our colleagues and students." (From the Preface.)

This volume is intended for the reader who wishes to obtain an overview of the variety of testing devices used in schools including teacher-made tests, standardized achievement tests, tests of special aptitudes, and tests used in counseling and guidance. The treatment of topics is generally nonstatistical although appendices are provided describing common statistical procedures. (Lathrop.)

See also: Item #6.

IV. Foreign Language Testing

30. Ayer, George W. "An Auditory Discrimination Test Based on Spanish." Modern Language Journal, XLIV (May 1960), 227-230.

31. Birznieks, Paul and Mechthild. "Written Production Tests for the Audio-Lingual Classroom." The German Quarterly, XXXIX (May 1966), 358-364.

32. Borglum, G.P. and H.L. Edsall. "The Departmental Examination as an Instrument of Standardization." Modern Language Journal, XXXVII (Oct. 1953), 273-286.

33. Brière, Eugène. "Testing the Control of Parts of Speech in FL Compositions." Language Learning, XIV, 1-2 (1964), 1-10.

34. Brooks, Nelson. "New Developments in Foreign Language Testing." Educational Horizons (Fall 1964), 21-25.

"Investigates new trends in foreign language testing, including tapes, test evaluation and speaking tests. Also comments on role of professional teacher in language testing and the positive effect of good testing upon learning."

35. ----, Chairman. "Report of the Committee on Tests." Pp. 49-56 in Foreign Language Teachers and Tests: 1954 Northeast Conference on the Teaching of Foreign Languages--Reports of the Working Committees. Ed. Hunter Kellenberger. Providence, Rhode Island: Brown Univ.

All Northeast Conference Reports (1954 to date) are available from Materials Center, Modern Language Association, 62 Fifth Avenue, New York 10011. Price: $3.00 per year.

36. ----, Chairman. "Report of the Committee on Tests." Pp. 64-68 in Culture, Literature, and Articulation: 1955 Northeast Conference on the Teaching of Foreign Languages--Reports of the Working Committees. Ed. Germaine Brée. New York: New York Univ.

Available from: See item #35.

37. ----, Chairman. "Report of the Committee on Tests: Spoken Language Tests." Pp. 20-23 in The Language Classroom: 1957 Northeast Conference on the Teaching of Foreign Languages--Reports of the Working Committees. Ed. William F. Bottiglia. Cambridge: Massachusetts Institute of Technology.

Available from: See item #35.

38. ----, Chairman. "Report of Working Committee IV: Definition of Language Competences Through Testing." Pp. 49-56 in The Language Learner: 1959 Northeast Conference on the Teaching of Foreign Languages--Reports of the Working Committees. Ed. Frederick D. Eddy. Boulder: Univ. of Colorado.

Available from: See item #35.

39. ---- "Tests and Measurements." Chapter 15 of Language and Language Learning: Theory and Practice. Second Edition. New York: Harcourt, Brace & World, 1964, pp. 199-225.

40. ---- "Making your Own Language Tests (Suggestions to teachers for preparing classroom examinations)." Modern Languages: Teaching and Testing. A work kit to accompany a film and filmstrip program. Developed by Cooperative Test Division, Educational Testing Service, Princeton, New Jersey, 1966 (pp. 9-21).

Materials developed with the assistance of the Modern Language Association of America to illustrate the audiolingual approach to teaching and to provide guidelines for making better classroom tests. The kit includes: 1 filmstrip, 1 long-playing record, 26 pre-printed Ditto-Masters. Complete set of materials: $10.

Available from: Cooperative Test Division, Educational Testing Service, Box 999, Princeton, New Jersey 08540.

41. Bryan, Miriam M. "The MLA Cooperative Foreign Language Tests: An Innovation in Foreign Language Testing." The Bulletin of the Pennsylvania State Modern Language Association, XLIII (April 1965), 97-100.

42. Carroll, John B. "Fundamental Considerations in Testing for English Language Proficiency of Foreign Students." Testing the English Proficiency of Foreign Students. (Reports of a conference sponsored by the Center for Applied Linguistics in cooperation with the Institute of International Education and the National Association of Foreign Student Advisors.) Washington, D.C.: Center for Applied Linguistics, 1961, pp. 30-40.

Reprinted in Teaching English as a Second Language: A Book of Readings. Ed. Harold B. Allen. New York: McGraw-Hill, 1965, pp. 364-372.

43. Clark, John L.D. "MLA Cooperative Foreign Language Tests." Journal of Educational Measurement, II (Dec. 1965), 234-244.

44. Deitz, Patricia. "An Oral Performance Rating Sheet." The French Review, XXXV (Oct. 1961), 54-58.

Contains very practical suggestions about oral testing.

45. Delattre, Pierre. "Comment tester la facilité de parole dans un laboratoire de langue." *Le français dans le monde*, 3 (août-septembre 1961), 36-38.

46. ---- "Testing Student's Progress in the Language Laboratory." *International Journal of American Linguistics*, XXVI (Oct. 1960), 77-98.

46a. *A Description of the College Board Supplementary Achievement Tests*. New York: College Entrance Examination Board, Revised annually. Approx. 60 pp.

Describes the following Supplementary Achievement Tests: French, German, Italian, Russian and Spanish Listening Comprehension Tests; Greek Achievement Test; Italian Achievement Test. The College Board offers secondary schools these tests each year.

Available free of charge: See item #11.

47. Dufort, Mary R. "An Audio-Comprehension Ability Test for FL Programs in the Elementary School." *California Journal of Elementary Education*, XXX (Nov. 1961), 121-128.

48. *English Testing Guidebook, Part I: Instructions for Rating Aural/Oral Proficiency in English Using the AULC Interview Rating Form*. By David P. Harris. Prepared at The American University Language Center for the International Cooperation Administration. 11 pp.

49. Eriksson, Marguerite, Ilse Forest, and Ruth Mulhauser. "Techniques of Review and Testing." Chapter VI of *Foreign Languages in the Elementary School*. Englewood Cliffs, N.J.: Prentice-Hall, 1964, pp. 53-69.

50. "Evaluation and Testing." *Bulletin of the California State Department of Education: French--Listening, Speaking, Reading, Writing*. Sacramento: California State Department of Education, 1962, pp. 115-130.

51. "Evaluation and Testing." *Bulletin of the California State Department of Education: Spanish--Listening, Speaking, Reading, Writing*. Sacramento: California State Department of Education, 1961, pp. 44-56.

52. Guerra, Emilio L., David A. Abramson, and Maxim Newmark. "The New York City Foreign Language Oral Ability Rating Scale." *The Modern Language Journal*, XLVIII (Dec. 1964), 486-489.

53. Hitchman, P.J. "The Testing of Spoken English: A Review of Research." *Educational Research*, VII (Nov. 1964), 55-72.

54. Jones, Willis Knapp. "A Prognosis Test for Spanish." *Hispania*, XLVIII, (March 1965), 104-109.

55. Kaulfers, Walter V. "Wartime Development in Modern-Language Achievement Testing." *The Modern Language Journal*, XXVIII (Feb. 1944), 136-150.

"The theory and practice of aural-comprehension and oral-fluency testing illustrated with concrete examples" (Author's summary). "The foregoing discussion is obviously not presented with the thought that the problems presented by oral fluency testing are

solved herewith, but rather to indicate possibilities and practical lines of approach to those who are interested in pioneering in a heretofore unexplored, but increasingly important field. The only real handicap to effective progress is the 'correlation fallacy' or common delusion that a high total-score on silent group-tests of vocabulary, grammar, or reading can automatically be taken to mean readiness to speak the language fluently in actual life-situations. In fact, not even ability to understand the spoken language provides any guarantee of comparable ability to speak it. One need only look at the world about one to find the proof in countless numbers" (pp. 149-150). A lot has happened in the field of foreign language testing since this article was written, but much remains to be done with respect to what Kaulfers calls "oral-fluency" testing. A comparison between the situation in 1944 and what is now being done in schools and colleges with respect to "oral-fluency" testing may be quite revealing.

56. Keesee, Elizabeth. "Testing Oral Skills." Modern Foreign Languages in the Elementary School: Teaching Technqiues. Washington, D.C.: U.S. Department of Health, Education, and Welfare, 1960, pp. 55-61.

57. Lado, Robert. "How to Test Cross-Cultural Understanding." Studies in Languages and Linguistics in Honor of Charles C. Fries. Ed. Albert H. Marckwardt. Ann Arbor: English Language Institute, Univ. of Michigan, 1964, pp. 353-362.

58. ---- "Language Testing." Chapter 16 of Language Teaching: A Scientific Approach. New York: McGraw-Hill, 1964, pp. 158-170.

59. ---- Language Testing: The Construction and Use of Foreign Language Tests. A Teacher's Book. London: Longmans, Green, 1961. xxiii, 389 pp. Bibliographical notes and selected references after some chapters.

The complete text of the London edition was published in the United States, without abridgment or change, by McGraw-Hill Book Company in 1964.

"Primarily intended for teachers of foreign languages and of English as a second language. Beginning with a discussion of the nature of language and language learning, it proceeds to a consideration of how the various language skills may be tested. Techniques for measuring cross-cultural understanding are also proposed. Several chapters at the end deal with fundamental test statistics." (David P.Harris.)

60. Language Structure at FLES Level Including Testing for Mastery of Structures. New York: National Information Bureau of AATF, 1962.

61. *MLA Bulletin of Information: 1967-68 MLA Foreign Language Proficiency Tests for Teachers and Advanced Students*. See item #6.

62. *MLA Cooperative Foreign Language Tests*. *Handbook*. 24 pp.
Directions for Administering and Scoring. 40 pp.
Booklet of Norms. 82 pp.
Princeton, N.J.: Educational Testing Service, Cooperative Test Division, 1965.

Available from: Cooperative Test Division, Educational Testing Service Box 999, Princeton, New Jersey. Price of each booklet: $1.00.

63. *MLA-Cooperative Foreign Language Tests . . . in Brief*. Princeton, N.J.: Cooperative Test Division, Educational Testing Service, 1964.

Available free of charge: See item #40.

Describes the MLA foreign language tests for secondary schools and colleges (French, German, Italian, Russian, Spanish). See item #62.

64. Mueller, Theodore. "Grade Standards in Foreign Language Instruction." *French Review*, XXXII (April 1959), 443-447.

65. "Oral Rating Form for Rating Language Proficiency in Speaking and Understanding English." Washington, D.C.: The American Language Institute, Georgetown Univ., 1960.

66. Pike, Eunice V. "A Test for Predicting Phonetic Ability." *Language Learning*, IX (1959), 35-41.

67. Pimsleur, Paul. "A French Speaking Proficiency Test." *French Review*, XXXIV (April 1961), 470-479.

"Reports on a standardized French Speaking test which attempts to go as far as possible in the direction of complete objectivity. Reviews past efforts and discusses the present test in detail."

68. ---- "Predicting Achievement in Foreign Language Learning." *International Journal of American Linguistics*, XXIX, 2, Part III (April 1963), 129-136.

"Describes an aptitude battery, consisting of several sub-tests, each of which measures some relatively well-defined characteristic related to foreign language learning. Includes linguistic analysis test and sound-symbol test among others.

69. ---- "Predicting Success in High School Foreign Language Courses." *Educational and Psychological Measurement*, XXIII (Summer 1963), 349-357.

70. ---- "Testing Language Aptitude." International Conference on Modern Foreign Language Teaching. (Papers and Reports of Groups and Committees, Preprints--Part 2), published by Paedagogische Arbeitsstelle and Sekretariat Paedagogisches Zentrum, Berlin, 1964, pp. 49-60.

71. ---- "Testing the Speaking Skill." International Conference on Modern Foreign Language Teaching. (Papers and Reports of Groups and Committees, Preprints--Part 2), published by Paedagogische Arbeitsstelle and Sekretariat Paedagogisches Zentrum, Berlin, 1964, pp. 61-74.

72. ---- "Testing Foreign Language Learning." Trends in Language Teaching. Ed. Albert Valdman. New York: McGraw-Hill, 1966, pp. 175-214.

72a. Rice, Frank A. "The Foreign Service Institute Tests Language Proficiency." Linguistic Reporter, I (May 1959), 4.

73. Sapon, Stanley M. "Report of the Committee on Tests." Pp. 33-38 in Foreign Language Tests and Techniques: 1956 Northeast Conference on the Teaching of Foreign Languages--Reports of the Working Committees. Ed. Margaret Gilman. Bryn Mawr, Penn.: Bryn Mawr College.

"Having completed its work on a test of auditory comprehension, the Committee continues its efforts this year with an intensive examination of the needs and problems involved in the area of 'speaking tests,' or tests of oral production" (p. 34).

Available from: See item #35.

74. Scheider, Rose M. "Evolution of the Listening Comprehension Test." College Board Review, XLVIII (Fall 1962), 24-28.

75. Seelye, H. Ned. "Field Notes on Cross-Cultural Testing." Language Learning, XVI, 1-2 (1966), 77-85.

76. Stack, Edward M. "Monitoring and Testing: Remedial and Evaluative Techniques." The Language Laboratory and Modern Language Teaching. New York: Oxford Univ. Press, 1966.

77. Strain, Jeris E. "Difficulties in Measuring Pronunciation Improvement." Language Learning, XIII, 3-4 (1963), 217-224.

78. Testing the English Proficiency of Foreign Students. (Report of a Conference sponsored by the Center for Applied Linguistics in cooperation with the Institute of International Education and the National Association of Foreign Student Advisors.) Washington, D.C.: Center for Applied Linguistics, 1961. iii, 103 pp., "Additional Background Materials."

See also item #42.

79. Upshur, John A. "Language Proficiency Testing and the Contrastive Analysis Dilemma." Language Learning, XII (Sept. 1962), 123-127.

80. ---- "Cross-Cultural Testing: What to Test." Language Learning, XVI, 3-4 (1966), 183-196.

81. Valette, Rebecca M. "Language Quizzes: A Means of Increasing Laboratory Effectiveness." Foreign Language Annals, I (Oct. 1967), 45-48.

82. ---- Modern Language Testing: A Handbook. New York: Harcourt, Brace & World, 1967. xvi, 200 pp., bibliography, index.

"It is clear that. . .new methods of teaching require new methods of evaluation. This handbook introduces the teacher to a diversity of testing techniques based on modern measurement theory; the book's emphasis, however, is on the classroom situation, theory being introduced only when it has a direct application for the teacher. Many examples have been given to help the teacher prepare tests that will effectively evaluate proficiency in the four fundamental skills. A special section also discusses literature tests." (From the Preface.)

83. ---- "Oral Objective Testing in the Classroom." German Quarterly, XXXVIII (March 1965), 179-187.

84. ---- "The Use of the Dictée in the French Language Classroom." Modern Language Journal, XLVIII (Nov. 1964), 431-434.

"In this article I examine the reliability of the dictée both as a testing technique and a teaching technique for beginning French courses taught by an audio-lingual method. In the experiment, 120 college students were divided into two groups which differed only in that the first group was given a daily dictation. The analysis of performance on a common final examination indicates that (a) the dictée constitutes a valid test of overall language skills only for students with little practice in dictation, and that (b) emphasis on dictation develops proficiency in that skill alone without leading to proportional proficiency in other aspects of language learning." Rebecca M. Valette, ML Abstracts, #851.

85. Vasu, Leon V. "Measurement of Fluency--Direct and Indirect." Review, US Army School (Spring 1961), 1-5.

86. Wilkins, George W., Jr., and E. Lee Hoffman. "The Use of Cognates in Testing Pronunciation." Language Learning, XIV, 1-2 (1965), 39-43.

Addenda to Section IV

87. Dufau, Micheline, Chairman. "From School to College: The Problem of Continuity." Pp. 102-108 in _Foreign Language Teaching: Challenges to the Profession. Reports of the Working Committees, Northeast Conference on the Teaching of Foreign Languages_, 1965. Ed. G. Reginald Bishop, Jr. New Brunswick, N.J.: Rutgers Univ. (Section 4: Placement Tests, Section 5: Choosing a Placement Test.)

    Available from: See item #35.

88. Sadnavitch, Joseph M. and W. James Popham. "Measurement of Spanish Achievement in the Elementary School." _Modern Language Journal_, XLVI (Nov. 1961), 297-299.

V. Statistics and Interpretation of Evaluation Results

A. General

89. Diederich, Paul B. _Short-cut Statistics for Teacher-made Tests._ 2nd ed. Evaluation and Advisory Service Series No. 5. Princeton, N.J.: Educational Testing Service, 1964. 37 pp.

    Explains and demonstrates simple, time-saving techniques for use by teachers.

    "A useful guide to teachers who wish to go one step beyond the intuitive approach to examining test results. The style is sympathetic to the teacher who has had no prior experience with statistics." (Lathrop.)

    Available free of charge from: See item #14.

* 90. Ferguson, George A. _Statistical Analysis in Psychology and Education_. New York: McGraw-Hill, 1959. vii, 347 pp., index, tables, glossary of symbols, bibliography. A new edition appeared in 1966.

    "The object of this book is to introduce students and research workers in psychology and education to the concepts and applications of statistics. Emphasis is placed on the analysis and interpretation of data resulting from the conduct of experiments . . . I have attempted not only to introduce the student to the practical technology of statistics but also to explain in a non-mathematical and frequently intuitive way the nature of statistical ideas." (From the Preface.)

* 91. Guilford, J.P. _Fundamental Statistics in Psychology and Education_. Fourth Edition. New York: McGraw-Hill, 1965.

"An intermediate level treatment of descriptive statistics and measurements concepts. The treatment presumes some mathematical background for the reader, at least through basic algebra. The treatment of correlation, reliability, predictive validity and standard scores are well written for the reader who has had a prior elementary exposure to these concepts."

* 92. McCollough, Celeste and Loche Van Atta. Introduction to Descriptive Statistics and Correlation: A Program for Self-Instruction. New York: McGraw-Hill, 1965. xii, 159 pp., index.

"This text presents the elements of descriptive statistics and correlation in a form suitable for self-instruction. It has been prepared in response to the request for a briefer introduction to statistics than is provided in our earlier programmed text, Statistical Concepts: A Program for Self-Instruction [McGraw-Hill, 1963] . . . The book is intended to provide an introduction to elementary statistical concepts for students of applied social science. Emphasis has been placed on interpretation of these concepts rather than on the computation of statistical measures or their mathematical derivation. Although the book can be of assistance to students taking a formal statistics course, it is not meant to serve as a textbook or a substitute for such a course." (From the Preface.)

B. Norms, Scores, and Their Interpretation

93. English Testing Guidebook, Part II: Fundamentals of Test Construction and Interpretation. See item #15.

94. Lyman, Howard B. Test Scores and What They Mean. Englewood Cliffs, N.J.: Prentice-Hall, 1963.

95. McLaughlin, Kenneth F. Interpretation of Test Results. Bulletin 1964, No. 7, Office of Education, U.S. Department of Health, Education and Welfare, Washington, D.C., 1964. vi, 63 pp., selected references.

Available from: U.S. Government Printing Office, Washington, D.C. Price: $0.30.

96. Wormer, Frank B. Test Norms: Their Use and Interpretation. Washington, D.C.: National Association of Secondary-School Principals, 1965.

Bibliographical Sources

97. Buros, Oscar Krisen, ed. The Sixth Mental Measurements Yearbook. Highland Park, N.J.: Gryphon Press, 1965. xxxvi, 1714 pp., indices.

    The best single reference source in the field of testing and measurement problems. Previous publications by Buros appeared in 1936, 1937, 1938, 1949, 1953 and 1959.

98. Locating Information on Educational Measurement: Sources and References. Evaluation and Advisory Service Series No. 1. Princeton, N.J.: Educational Testing Service, 1965. 39 pp.

    Provides suggestions for using various sources of measurement information, including reference volumes, books, professional journals, and test publishers. An excellent source of information.

    Available free of charge: See item #14.

99. Nostrand, Howard Lee, David William Foster, and Clay Benjamin Christensen. Research on Language Teaching: An Annotated Bibliography, 1945-64. Second edition, revised. Seattle: Univ. of Washington Press, 1965. xxi, 373 pp.

    Section 2.7 Evaluation of Achievement in Language Learning, pp. 173-184.
    Section 3.3 Aptitutde and Prognostic Tests, pp. 216-226.